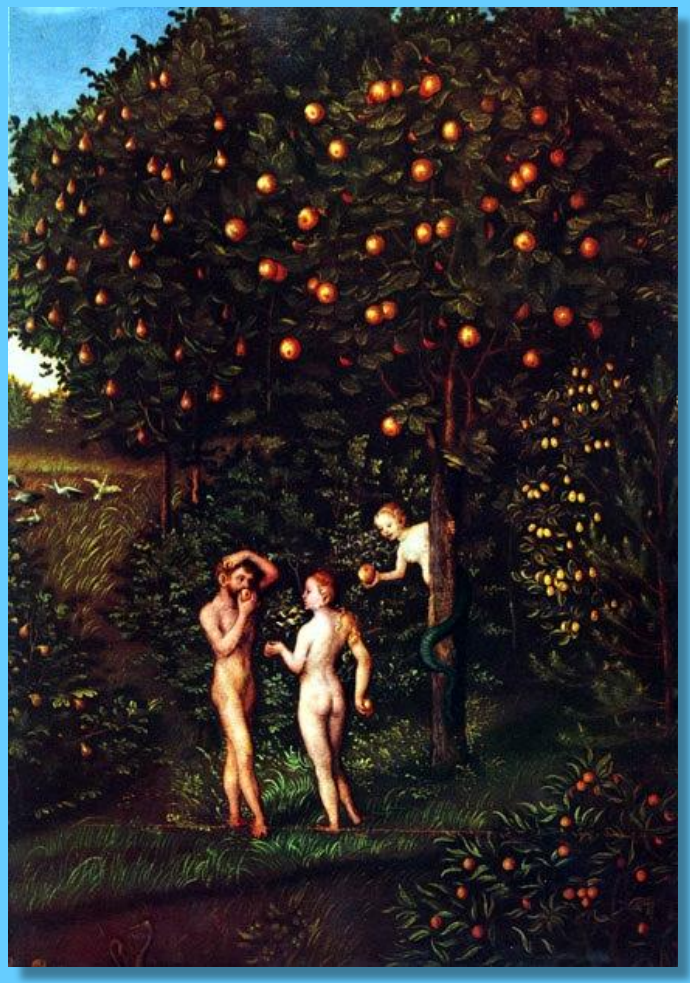
Statistics: Testing Hypotheses; the critical ratio.

- ✓ Click “slide show” to start this presentation as a show.
- ✓ **Remember: focus & think about each point; do not just passively click.**

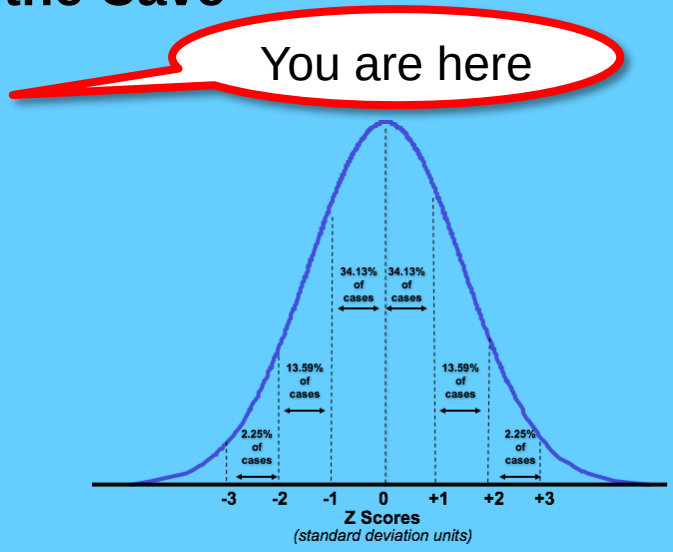


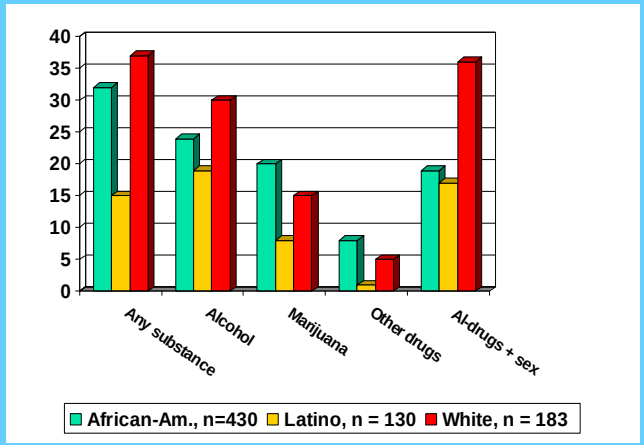
Cranach, *Tree of Knowledge* [of Good and Evil] (1472)

© Dr. David J. McKirnan, 2014
The University of Illinois Chicago
McKirnanUIC@gmail.com
Do not use or reproduce without permission

SHOUT
SHARE
IT

1. Introduction to statistics & number scales
2. The Z score and the normal distribution
3. The logic of research; Plato's Allegory of the Cave
4. Testing hypotheses: The critical ratio
5. Calculating a t score
6. Testing t: The Central Limit Theorem
7. Correlations: Measures of association





© Dr. David J. McKirnan, 2014
The University of Illinois Chicago
McKirnanUIC@gmail.com
Do not use or reproduce without permission

Evaluating data

Here we will see how to use Z scores to evaluate data, and will introduce the concept of critical ratio.

- ☑ **Using Z scores to evaluate data**
- ❑ **Testing hypotheses: the critical ratio.**

Using Z scores

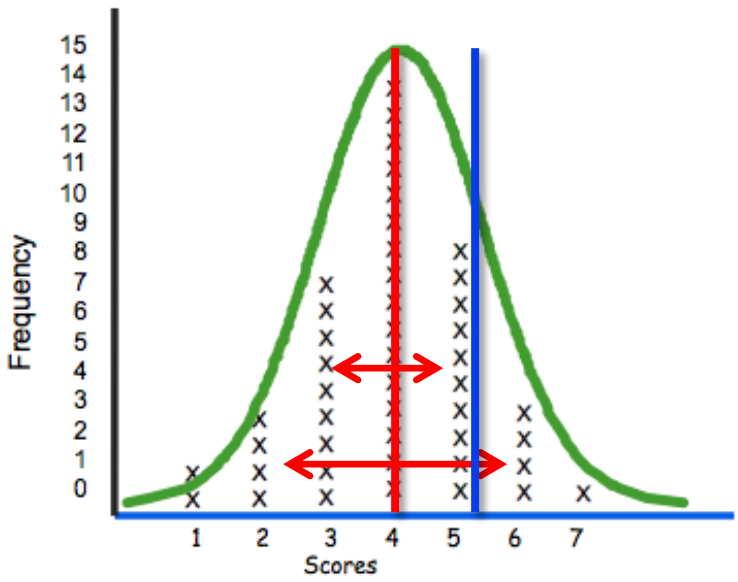
Modules 2 & 3 introduced key statistical concepts:

- Individual scores (**X**) on a variable,
- The Mean (**M**) of a set of scores,
- The Standard Deviation (**S**), reflecting the variance of scores around that Mean or average,
- The **Z** score; a measure of how far a score is above or below the Mean, divided by the Standard deviation:

$$Z = \frac{X - M}{S}$$

- Z is a basic form of **Critical Ratio**
- Now we will talk about using the critical ratio in statistical decision making.

Using Z to evaluate data



Z is at the core of how we use statistics to evaluate data.

Z indicates how far a score is from the M relative to the other scores in the sample.

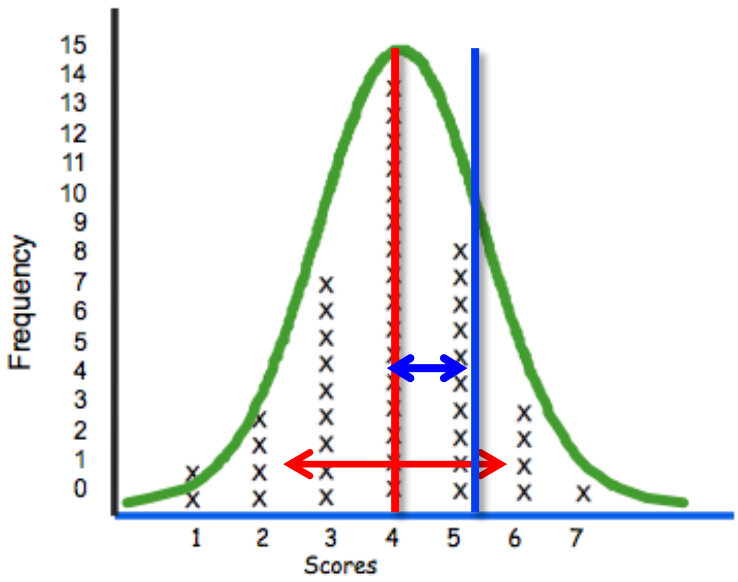
Z combines…

- ✓ **A score**
- ✓ **The M of all scores in the sample**
- ✓ **The variance in scores above and below M.**

Using Z to evaluate data

Z is at the core of how we use statistics to evaluate data.

Z indicates how far a score is from the M **relative to** the other scores in the sample.



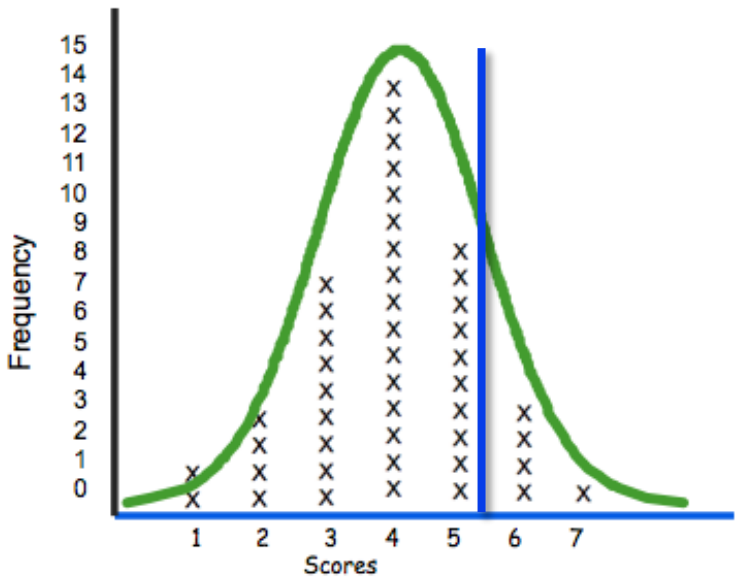
So…

- ✓ **If X** (an observed score) **= 5.2**
- ✓ **And M** (The Mean score) **= 4**

} **X - M = 1.2**

- ✓ **If S** (Standard deviation of all scores in the sample) **= 1.15**

Using Z to evaluate data



Z is at the core of how we use statistics to evaluate data.

Z indicates how far a score is from the M **relative to** the other scores in the sample.

So…

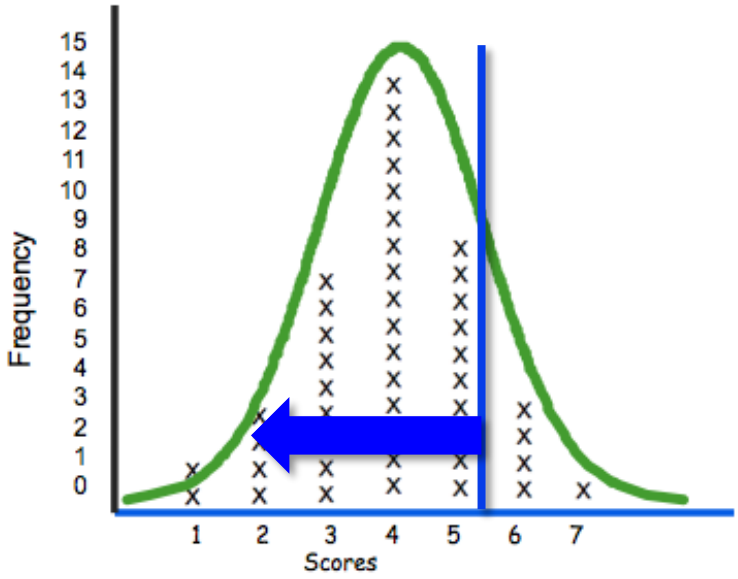
- ✓ **If X = 5.2**
- ✓ **And M = 4**

} **X - M = 1.2**

- ✓ **If S = 1.15 Z for our score is 1** (+)**.**

$$Z = \frac{X - M}{S} = \frac{5.2 - 4}{1.15} = \frac{1.2}{1.15} = 1.05$$

Using Z to evaluate data



Z is at the core of how we use statistics to evaluate data.

Z indicates how far a score is from the M **relative to** the other scores in the sample.

So…

- ✓ **If X = 5.2**
- ✓ **And M = 4**

} **X - M = 1.2**

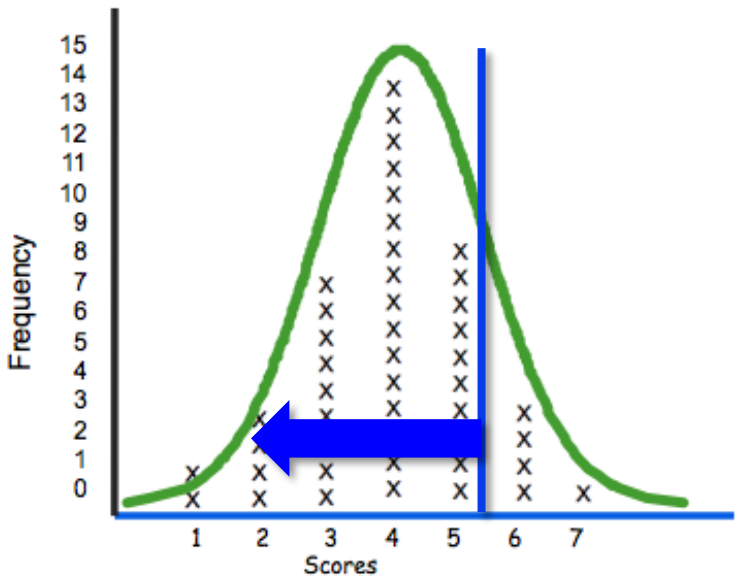
- ✓ **If S = 1.15 Z for our score is 1** (+)**.**
- ✓ **This tells us that our score is higher than ~ 84% of the other scores in the distribution.**

Using Z to evaluate data

Z is at the core of how we use statistics to evaluate data.

Z indicates how far a score is from the M **relative to** the other scores in the sample.

- ✓ This tells us that our score is higher than ~ 84% of the other scores in the distribution.



- ✓ **Unlike simple measurement with a ratio scale where a value – e.g. < 32º – has an absolute meaning.**
- ✓ **…inferential statistics evaluates a score relative to a distribution of scores.**

Z scores: areas under the normal curve, 2

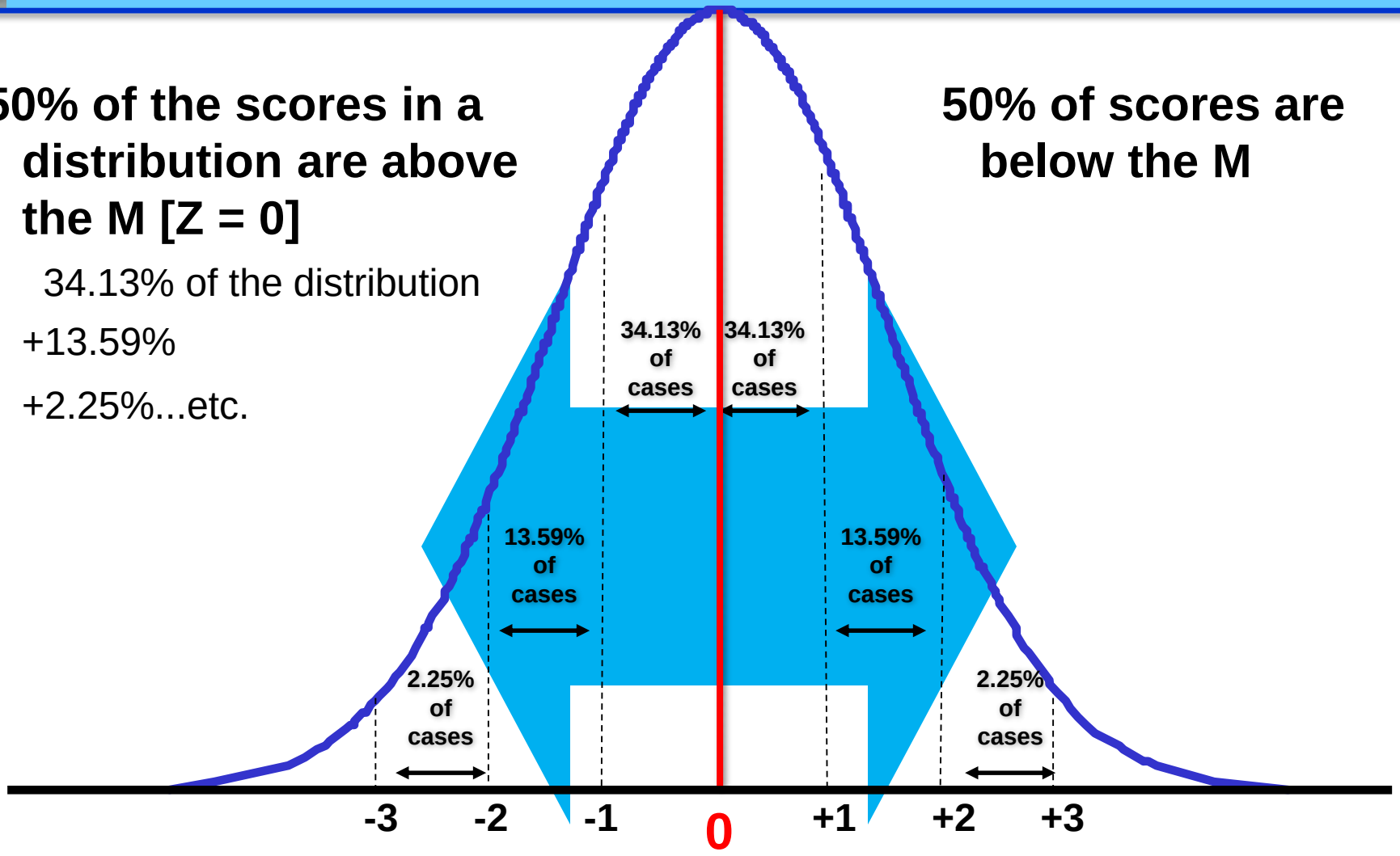
50% of the scores in a distribution are above the M [Z = 0]

34.13% of the distribution

+13.59%

+2.25%...etc.

50% of scores are below the M



Z Scores

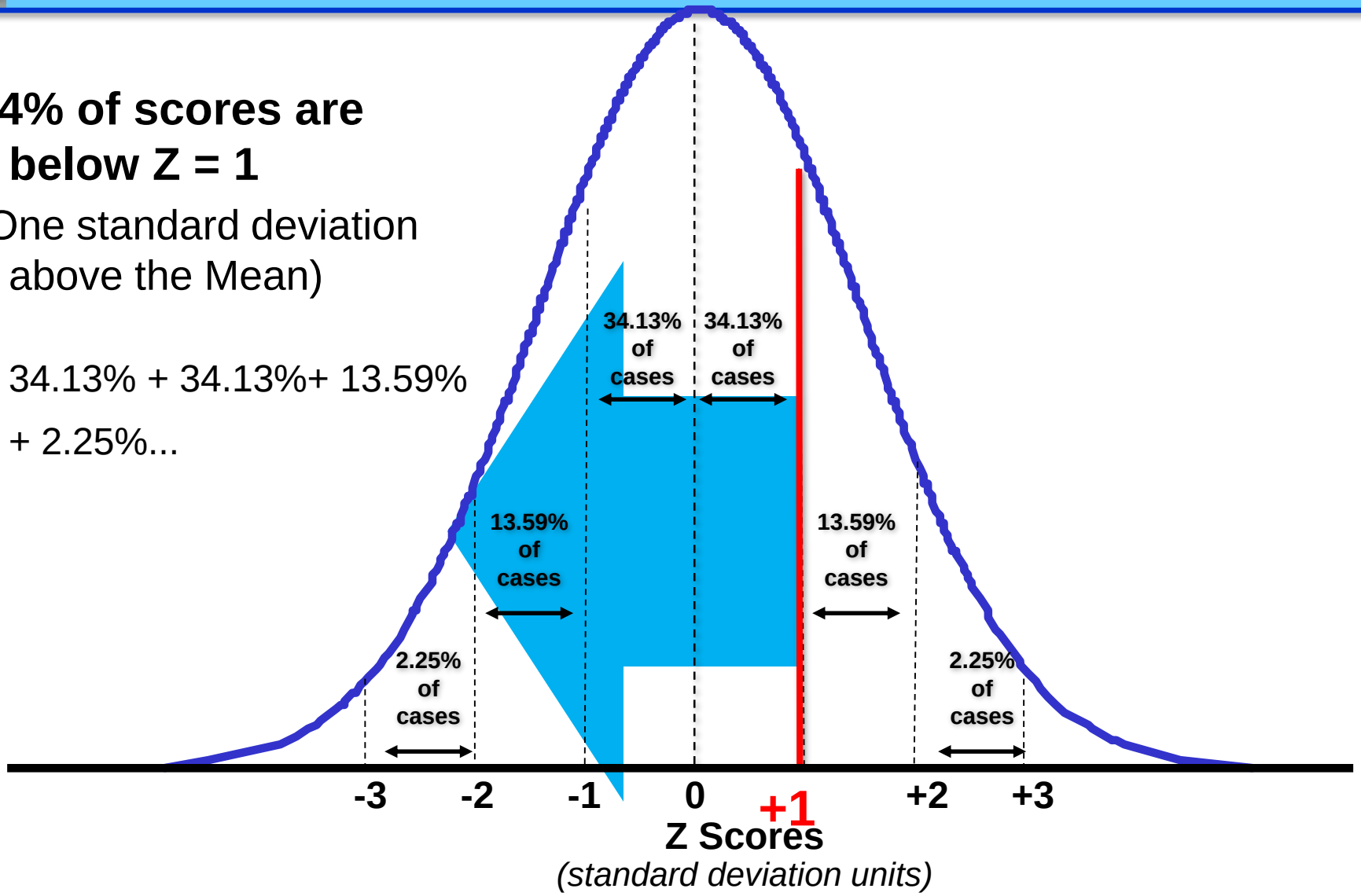
(standard deviation units)

Z scores: areas under the normal curve, 2

84% of scores are below Z = 1

(One standard deviation above the Mean)

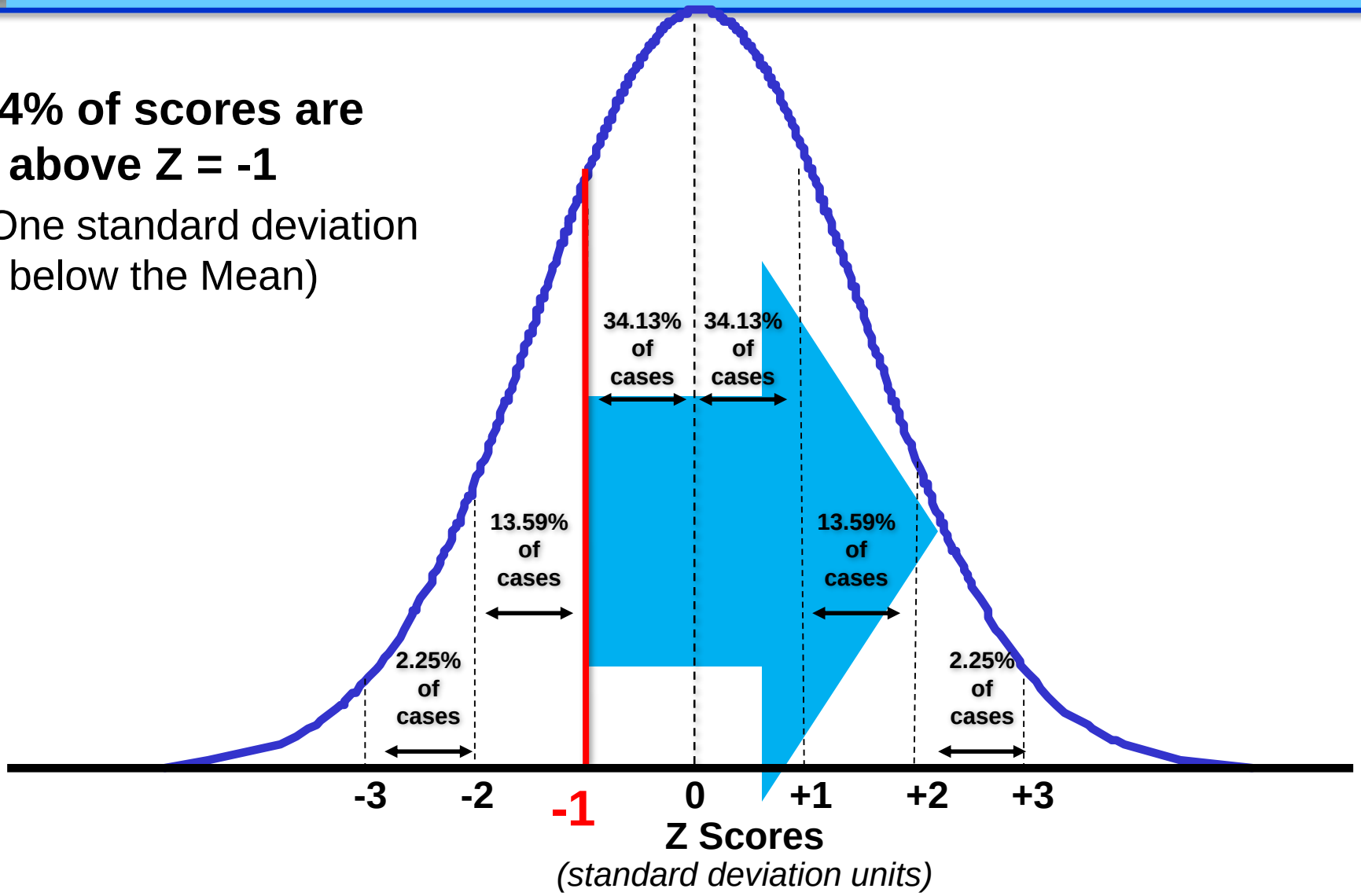
34.13% + 34.13%+ 13.59% + 2.25%...



Z scores: areas under the normal curve, 2

84% of scores are above Z = -1

(One standard deviation below the Mean)

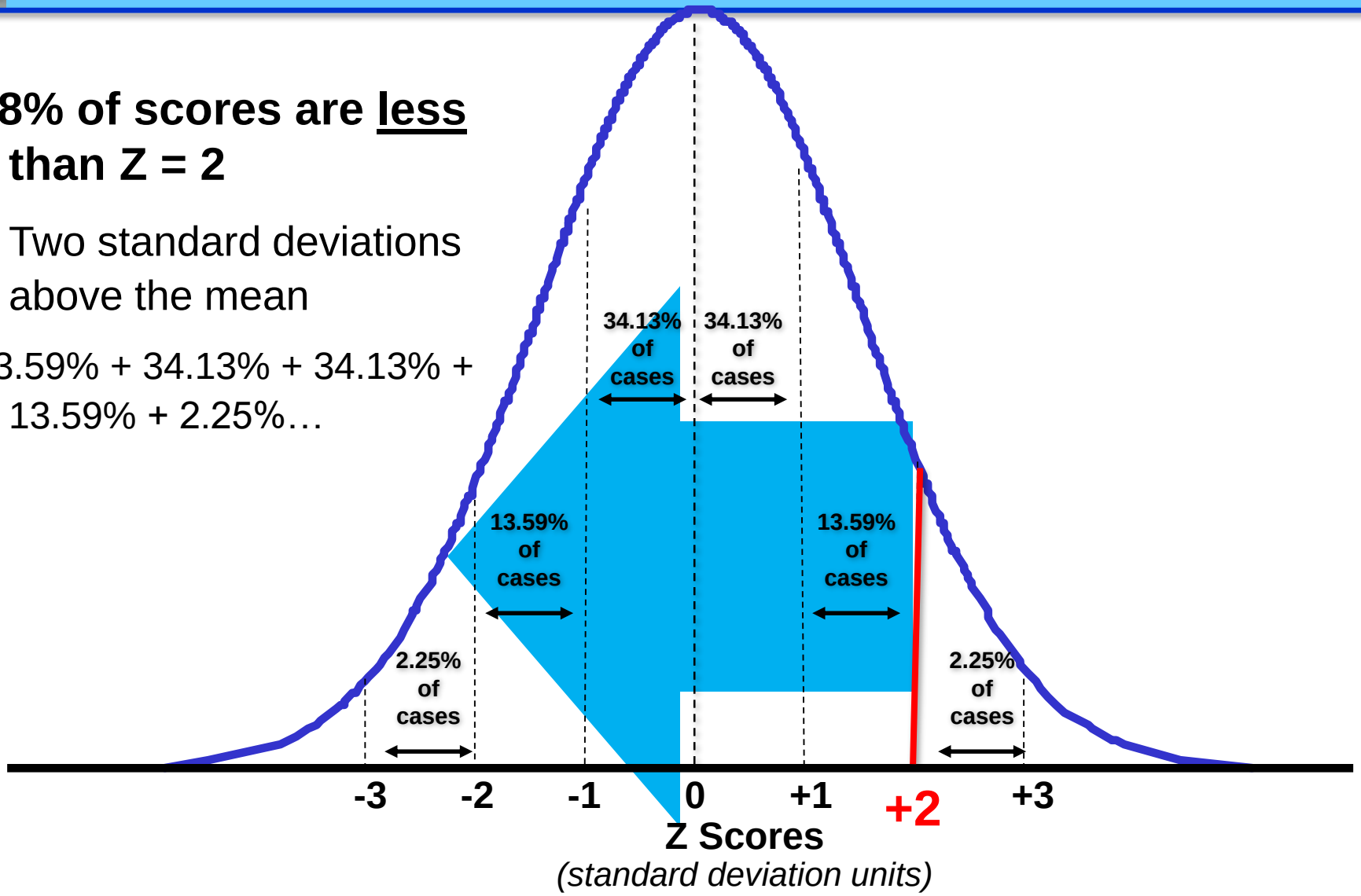


Z scores: areas under the normal curve, 2

98% of scores are <u>less</u> than Z = 2

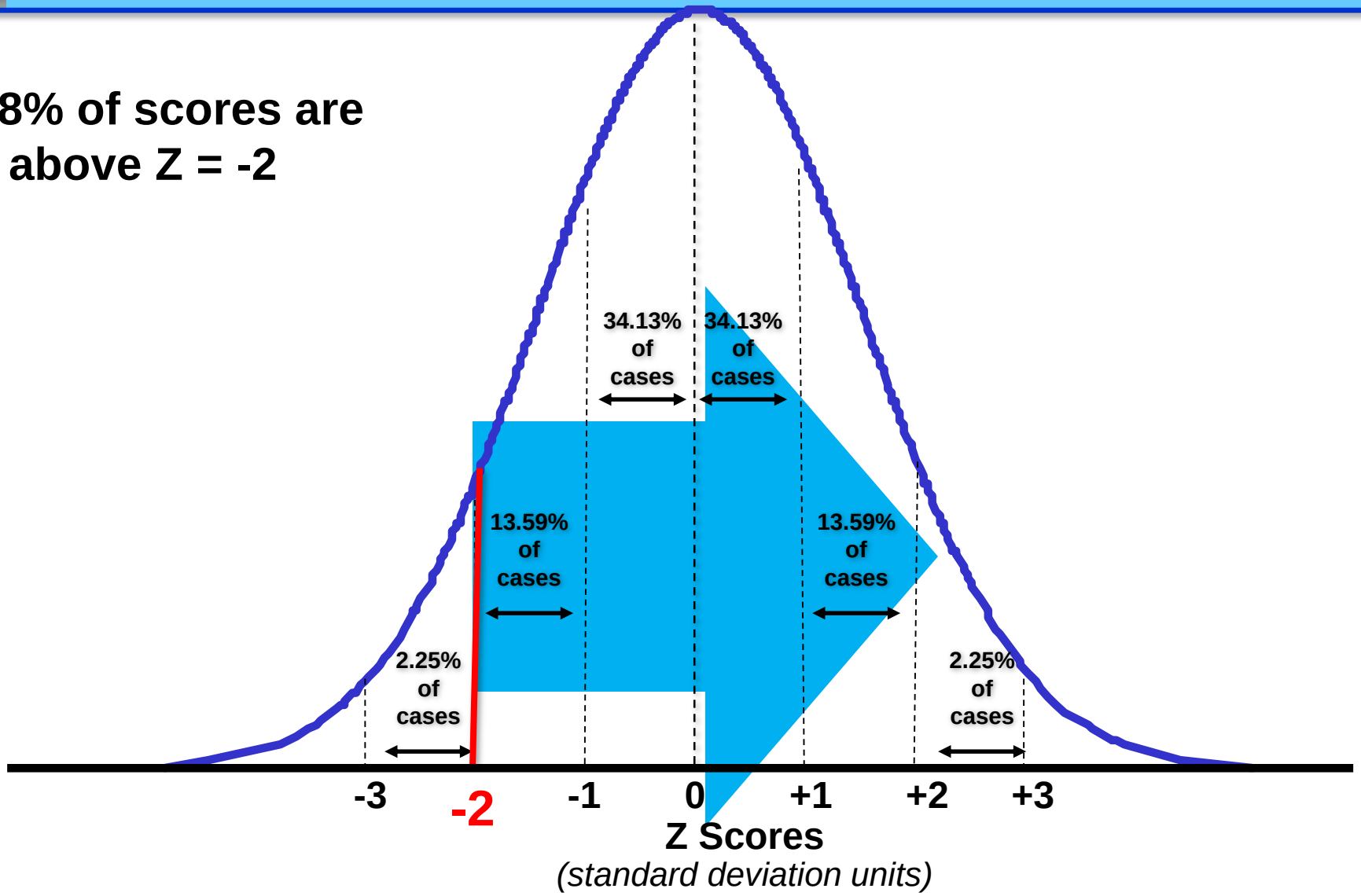
Two standard deviations above the mean

13.59% + 34.13% + 34.13% + 13.59% + 2.25%…

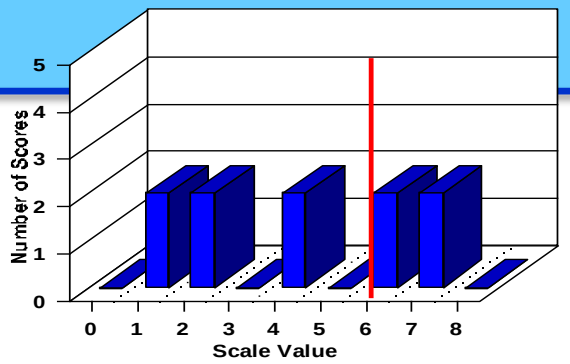


Z scores: areas under the normal curve, 2

98% of scores are above Z = -2

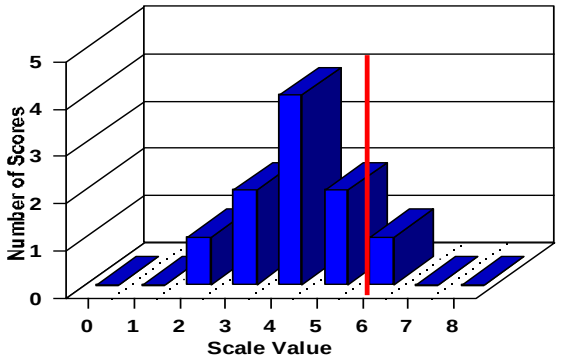


Evaluating Individual Scores



How good is a score of ‘6' in the group described in…

- **Table 1?**
- **Table 2?**



Evaluate in terms of:

A. The distance of the score from the M.

B. The variance in the rest of the sample

C. Your criterion for a “significantly good” score

A. The distance of the score from the *M*.

The participant is 2 units above the mean in both tables.

B. The variance in the rest of the sample:

Since Table 1 has more variance, a given score is not as good relative to the rest of the scores.

Table 1, high variance

$X - M = 6 - 4 = 2$

Standard Deviation $(S) = 2.4$.

$$Z = \frac{X - M}{S} = \frac{2}{2.4} = 0.88$$

About 70% of participants are below this Z score

Table 2, low(er) variance

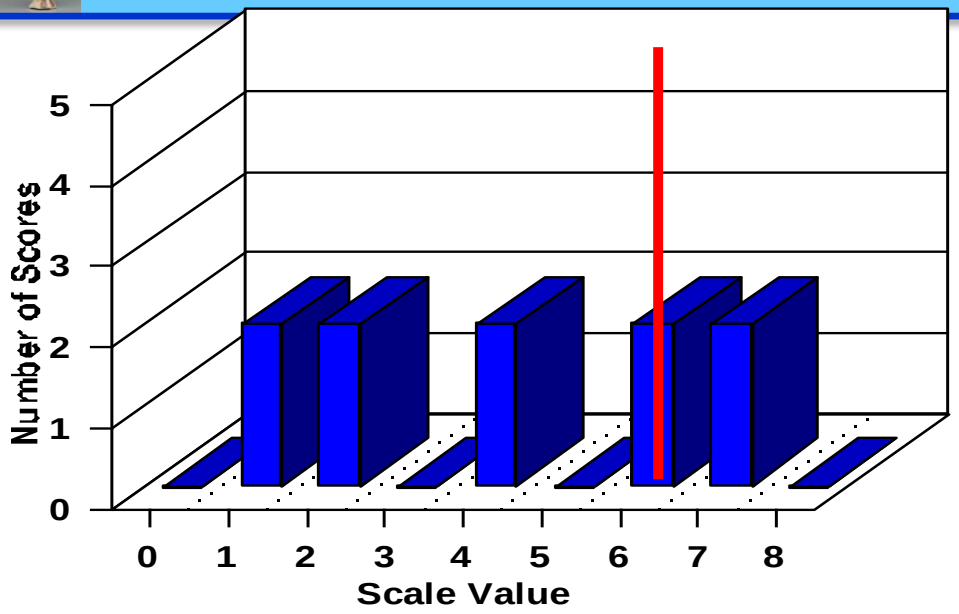
$X - M = 6 - 4 = 2$

Standard Deviation $(S) = 1.15$.

$$Z = \frac{2}{1.15} = 1.74$$

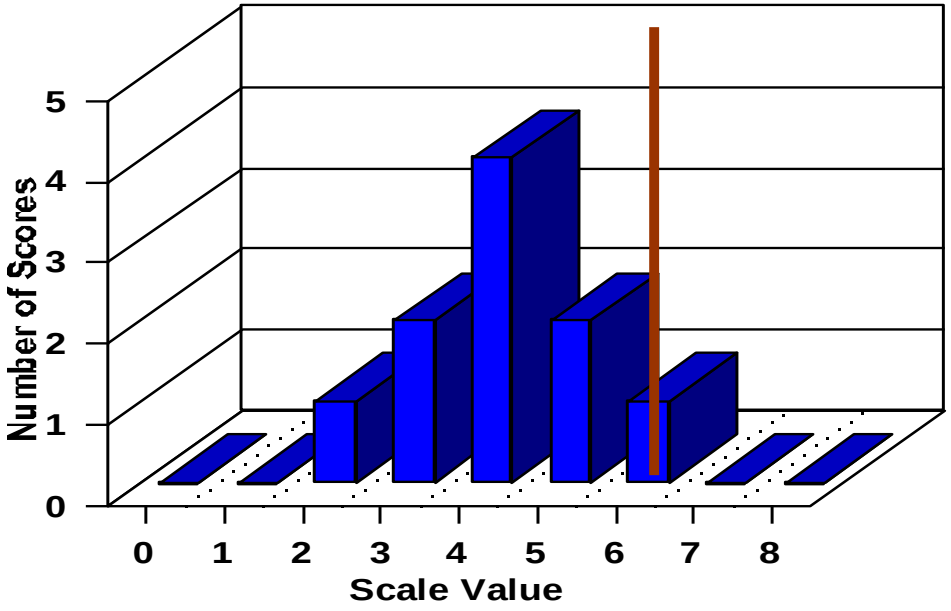
About 90% of participants are below this Z score

Comparing Scores: deviation x Variance



✓High variance
(S = 2.4)

'6' is not *that* high compared to rest of the distribution



✓ Less variance
(S = 1.15)

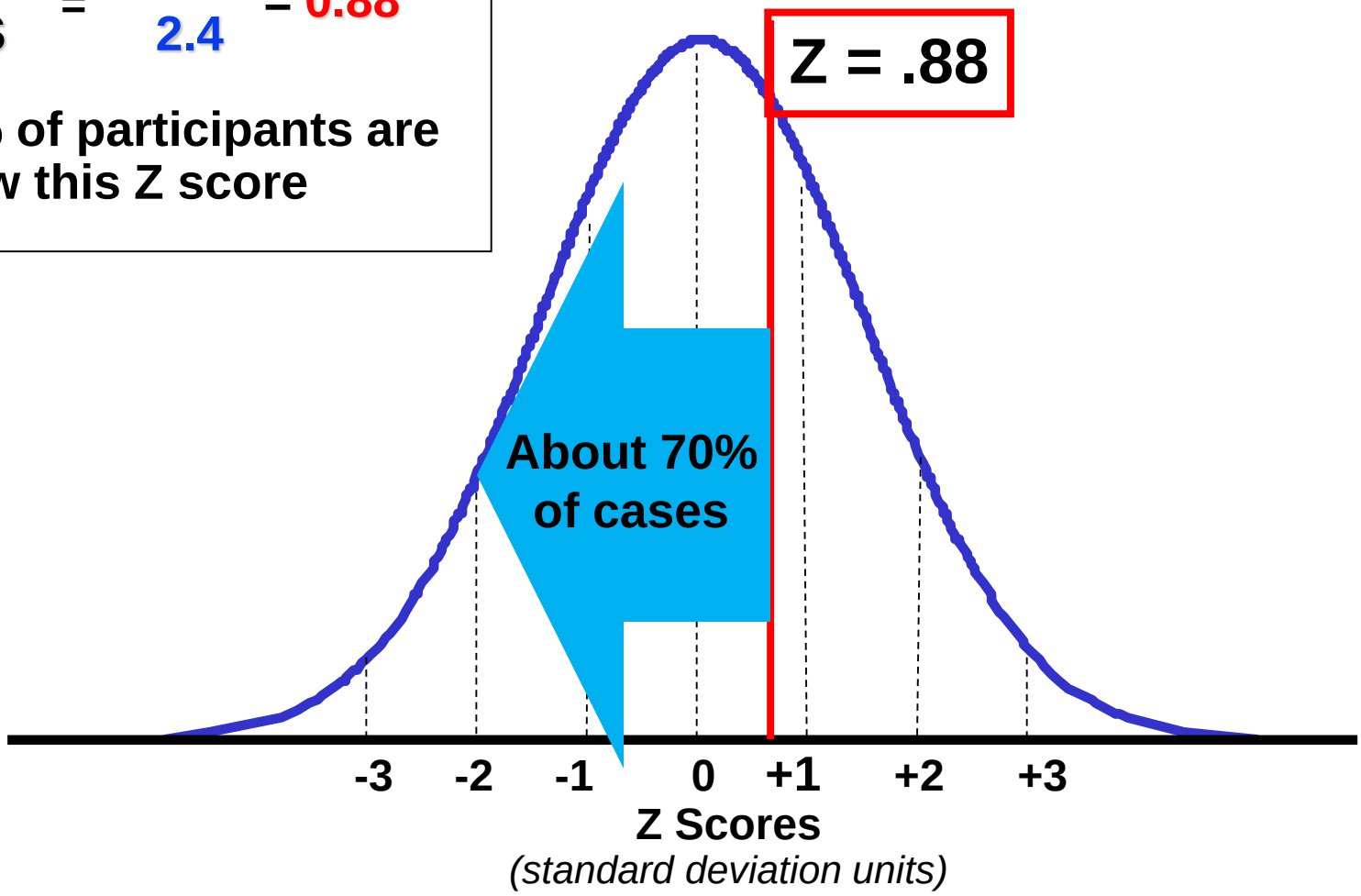
Here '6' is the highest score in the distribution

Normal distribution; high variance

Table 1, high variance

$$Z = \frac{X - M}{S} = \frac{2}{2.4} = 0.88$$

About 70% of participants are below this Z score

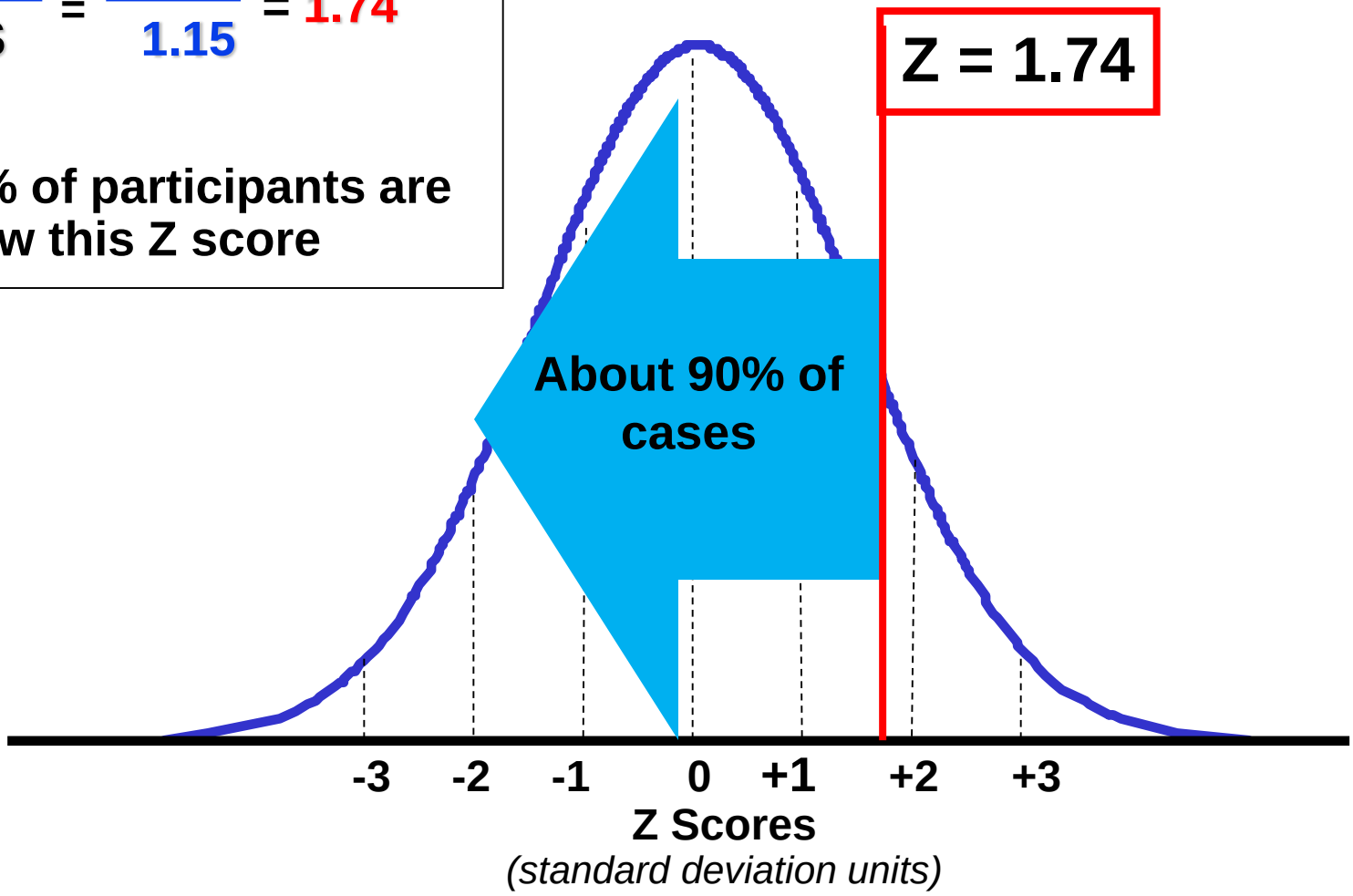


Normal distribution; low variance

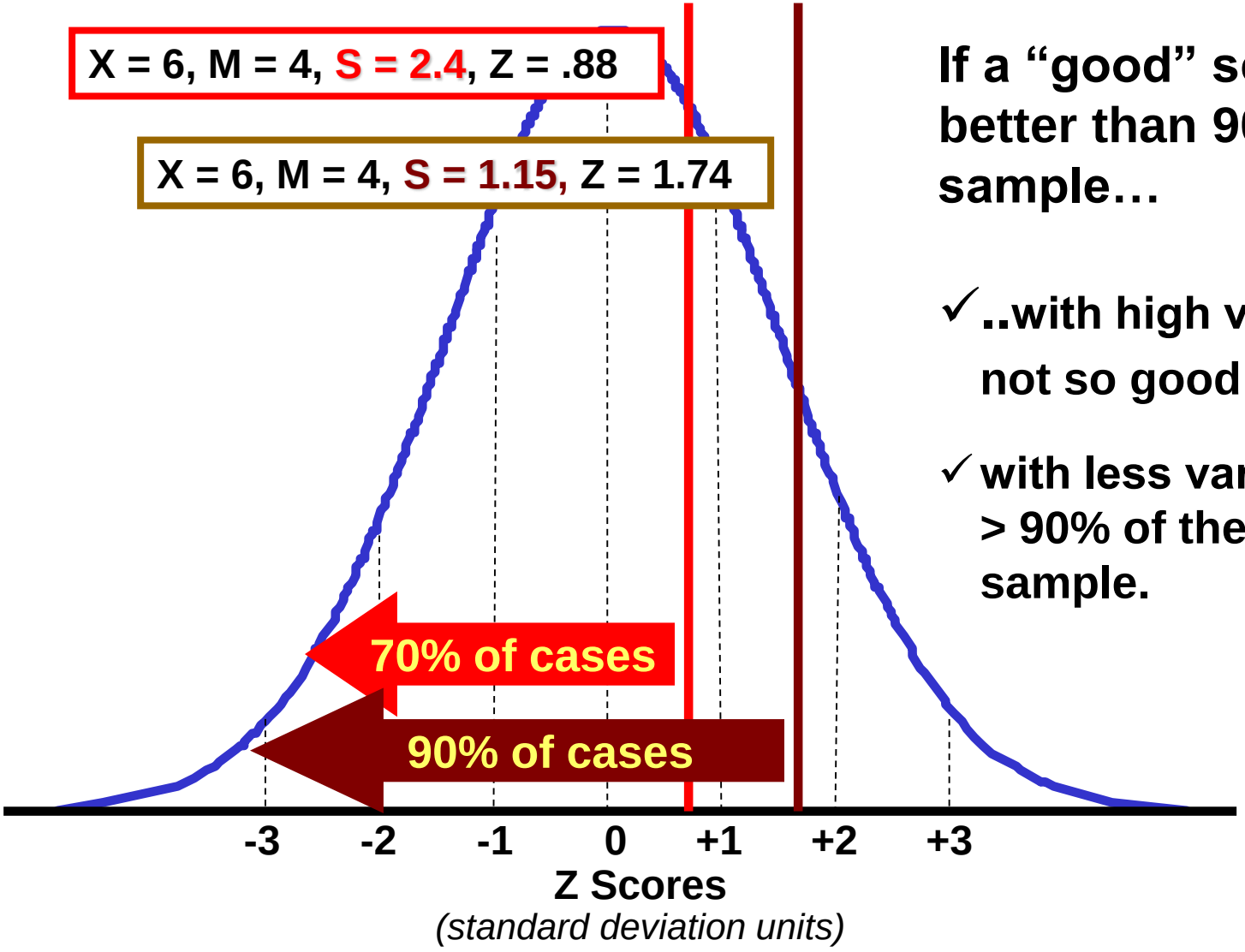
Table 2, low(er) variance

$$Z = \frac{X - M}{s} = \frac{2}{1.15} = 1.74$$

About 90% of participants are below this Z score



C. Criterion for a "significantly good" score



If a "good" score is better than 90% of the sample…

- ✓ **..with high variance '6' is not so good,**
- ✓ **with less variance '6' is > 90% of the rest of the sample.**

Summary: evaluating individual scores

How "good" is a score of '6' in two groups?

A. The distance of the score from the M.

In both groups X = 6 & M = 4, X – M = 2.

B. The variance in the rest of the sample

One group has low variance and one has higher.

*C. C*riterion for *"significantly good"* score

What % of the sample must the score be higher than…

Using Z to standardize scores

- **Z scores (or standard deviation units) *standardize* scores by putting them on a common scale.**
  - We want to combine two measures of social distance between racial groups.
  - For one, we measure how far someone stands from a member of a different group, ranging from 0 to 36 inches.
  - For the other, we give an attitude scale ranging from 1 ("Not distant at all") to 9 ("Very distant").
  - We cannot simply combine these measures; one has much higher raw scale values (ranging to 36) than the other (up to 9)…
  - Meaning they have very different Means and Standard Deviations.
  - To combine them we must put them on the same scale; we can change the raw values to Z scores, that will each have M = 0 and S = 1.
- **Any scores can be translated into Z scores for comparison…**

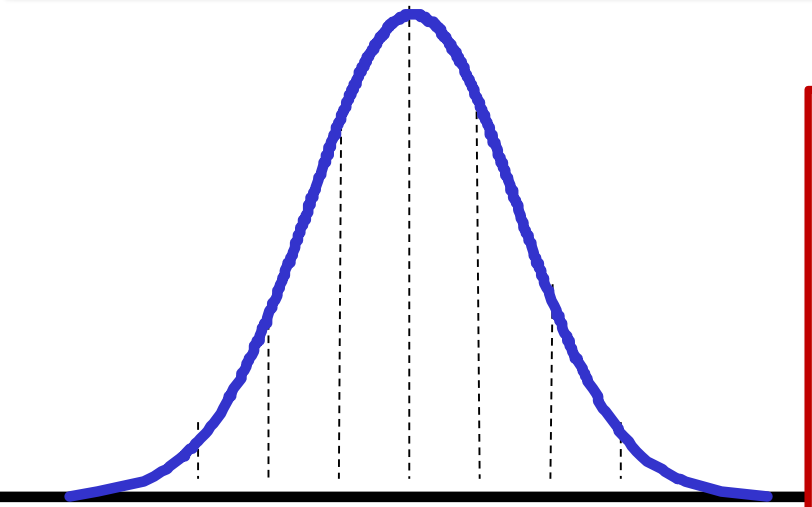
Using Z to standardize scores, cont.

- **Which is "faster"; a 2:03:00 marathon,**
- **or a 4 minute mile?**

Roberto Caucino / Shutterstock.com

Gustavo Miguel Fernandes / Shutterstock.com

- **We cannot directly compare these scores because they are on *different scales*.**
  - **One is measured in hours & minutes, one in 10^{ths} of a second.**
- **We can use Z scores to change each scale to common metric, where M = 0 and S = 1.**
- **Z scores can be compared, since they are *standardized* by being relative to the larger population of scores.**



2:50 2:45 2:40 2:30 2:20 2:15 2:10

Marathon times *(raw scores)*

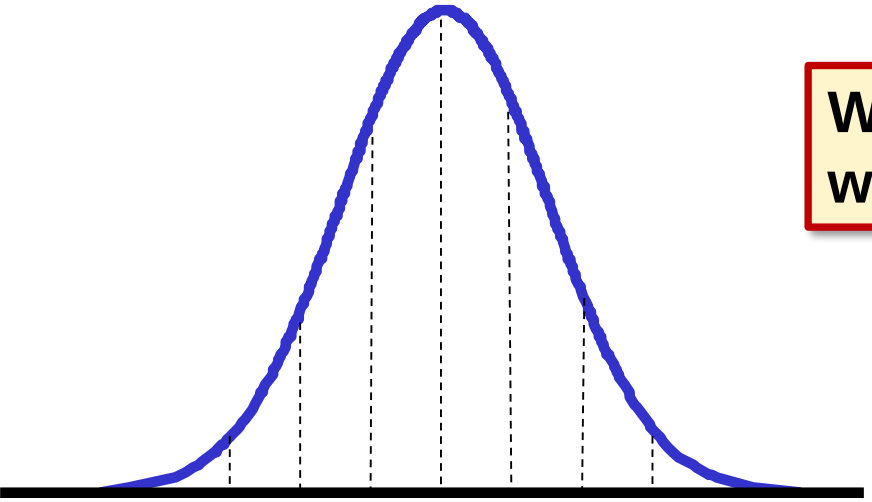
-4 -3 -2 -1 0 +1 +2 +3 +4

Z Scores *(standard deviation units)*

Here is a (made up) **distribution of world class marathon times.**

Using the Z formula, we can turn the raw scores into Z scores, that have M = 0 and S = 1.

Comparing Zs



4:30 4:25 4:20 4:10 4:00 3:50 3:45

Mile times

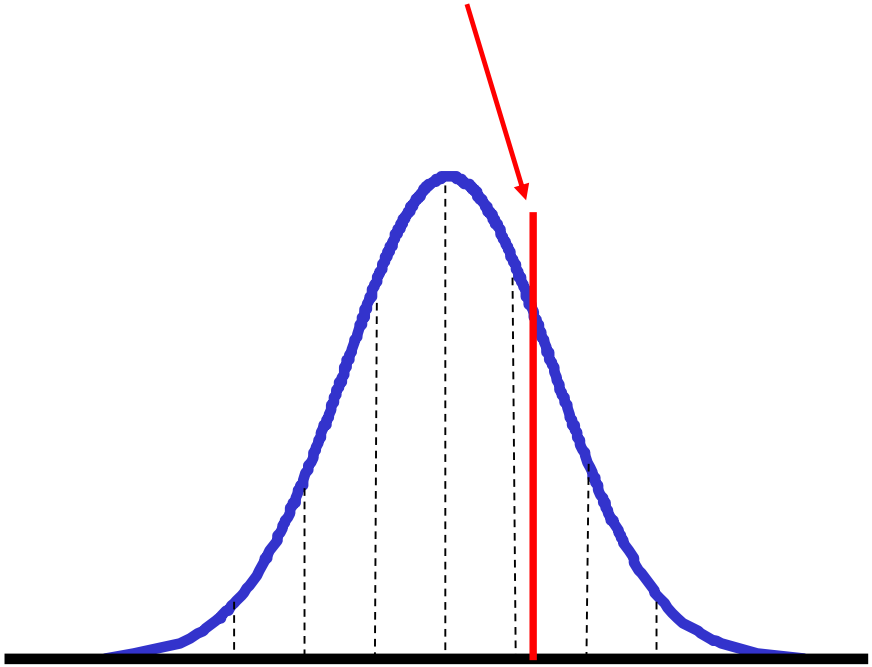
-3 -2 -1 0 +1 +2 +3

Z Scores ***(standard deviation units)***

We can do the same thing with our mile times.

By putting them on the same scale – Z scores (Standard Deviations) **– we can directly compare Marathon v. Mile times.**

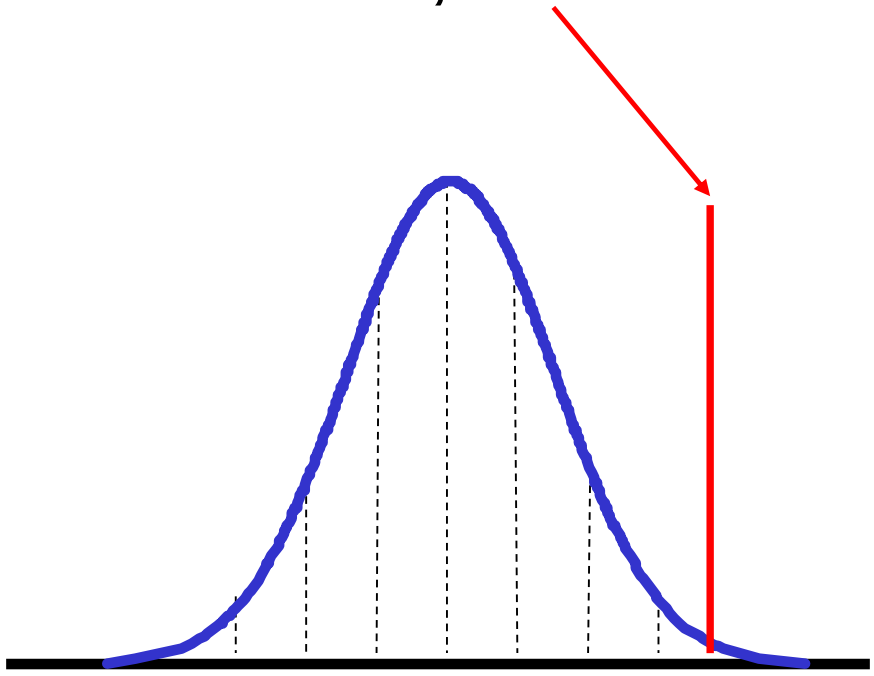
A 4 minute mile is not extreme: lots of people run faster. It translates to Z = 1.



4:30 4:25 4:20 4:10 4:00 3:50 3:45
Mile times

-3 -2 -1 0 +1 +2 +3
Z Scores ***(standard deviation units)***

A 2:03 marathon is very fast: only a few people have run it that fast. Here, Z = 4.

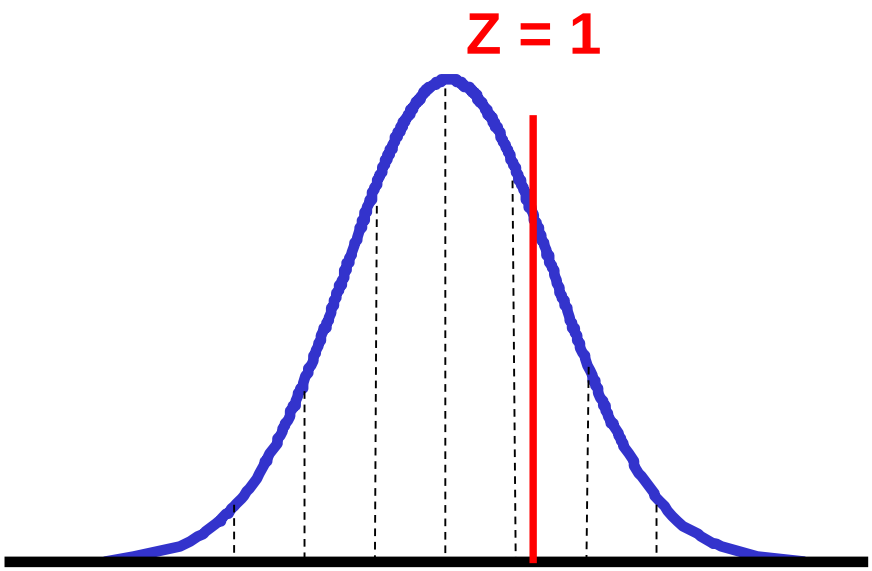


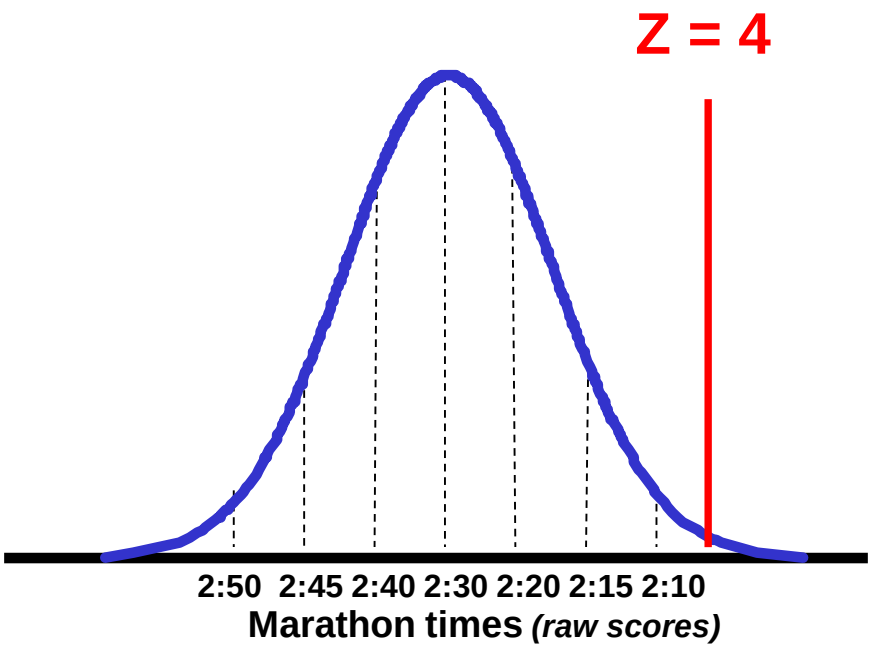
2:50 2:45 2:40 2:30 2:20 2:15 2:10
Marathon times ***(raw scores)***

-4 -3 -2 -1 0 +1 +2 +3 +4
Z Scores ***(standard deviation units)***

Standardizing Marathon & Mile times allows us to directly compare them…

We simply compare their Z scores to find that a 2:03 marathon is a good deal "faster" than a 4 minute mile.





The critical ratio

- **Using Z scores to evaluate data**
- **Testing hypotheses: the critical ratio.**

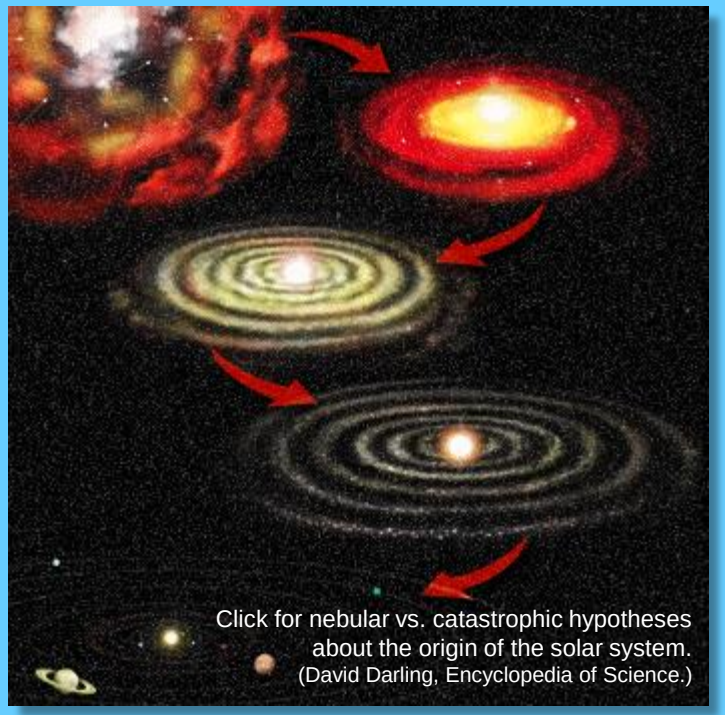

Click for nebular vs. catastrophic hypotheses about the origin of the solar system.
(David Darling, Encyclopedia of Science.)

Illustration of the nebular hypothesis

Using statistics to test hypotheses:

Core concept:

- ✓ **No scientific finding is "absolutely" true.**
- ✓ **Any effect is *probabilistic*:**
  - ➢ **We use empirical data to *infer* how the world words**
  - ➢ **We evaluate inferences by how likely the effect would be to occur by chance.**
  - ➢ **We use the normal distribution to help us determine how likely an experimental outcome would be by chance alone.**

Null Hypothesis: **All scores differ from the *M* by chance alone.**

Scientific observations are "innocent until proven guilty".

If we compare two groups or test how far a score is from the mean, the odds of their being different by chance alone is always greater than 0.

We cannot just take any result and call it meaningful, since any result may be due to chance, not the Independent Variable.

So, we assume any result is by chance unless it is strong enough to be unlikely to occur randomly.

Null Hypothesis: All scores differ from the *M* by chance alone.

Alternate (experimental) hypothesis: This score differs from M by more than we would expect by chance…

Using the Normal Distribution:

- More extreme scores have a lower probability of occurring by chance alone
- Z = the % of cases above or below the observed score
- A high Z score may be "extreme" enough for us to reject the null hypothesis

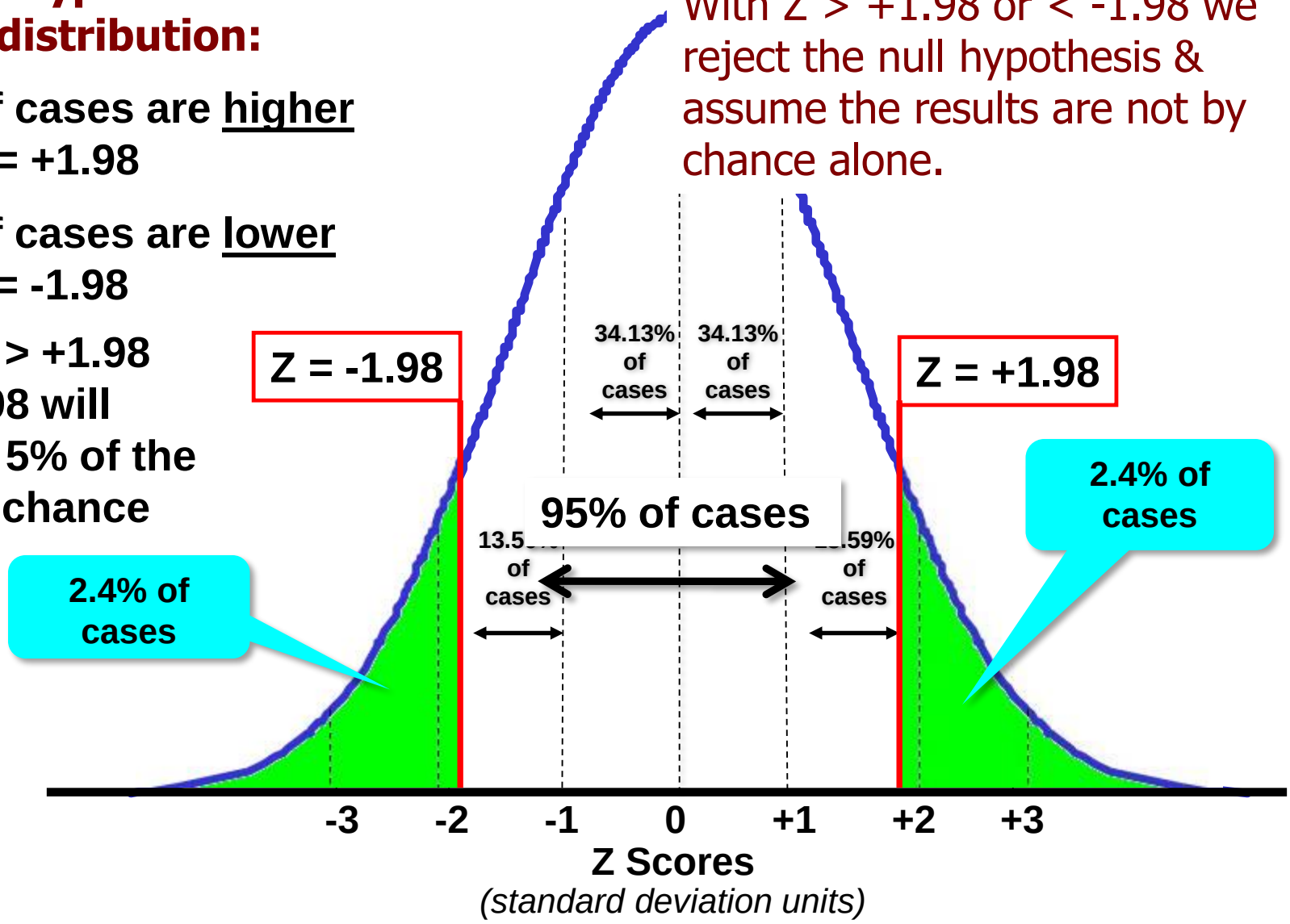
Statistical Significance

- **We assume a score with less than 5% probability of occurring**
  (i.e., higher or lower than 95% of the other scores… $p < .05$)
  is not by chance alone
- $Z > \pm 1.98$ **occurs < 95% of the time (**$p < .05$**).**
- **If** $Z > 1.98$ **we consider the score to be "significantly" different from the mean**
- **To test if an effect is "statistically significant"**
  - **Compute a Z score for the effect**
  - **Compare it to the critical value for p<.05;** $\pm$ **1.98**

In a hypothetical distribution:

- ✓ 2.4% of cases are higher than Z = +1.98
- ✓ 2.4% of cases are lower than Z = -1.98
- ✓ Thus, Z > +1.98 or < -1.98 will occur < 5% of the time by chance alone.

With Z > +1.98 or < -1.98 we reject the null hypothesis & assume the results are not by chance alone.



Evaluating Research Questions

Data	Statistical Question
One participant's score	Does this score differ from the M for the group *by more than chance?*
The mean for a group	Does this M differ from the M for the general population *by more than chance?*
Means for 2 or more groups	Is the difference between these Means *more than we would expect by chance*? -- more than the M difference between any 2 randomly selected groups?
Scores on two measured variables	Is the correlation ('r') between these variables *more than we would expect by chance* -- more than between any two randomly selected variables?

- ✓ To estimate the influence of chance we weight our results by the overall amount of variance in the data.
- ✓ In "noisy" data (a lot of error variance) we need a very strong result to conclude that it was unlikely to have occurred by chance alone.
- ✓ In very "clean" data (low variance) even a weak result may be statistically significant.
- ✓ This is the Critical Ratio:

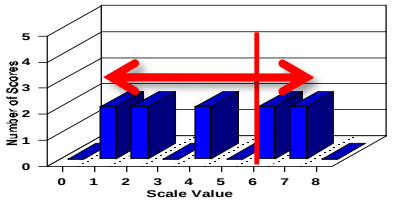
$$\text{Critical ratio} = \frac{\text{The strength of the results (experimental effect)}}{\text{Amount of error variance ("noise" in the data)}}$$

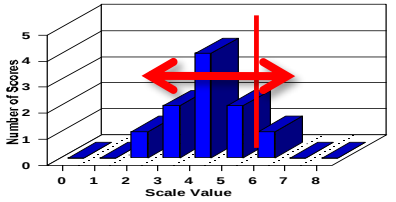
Critical ratio

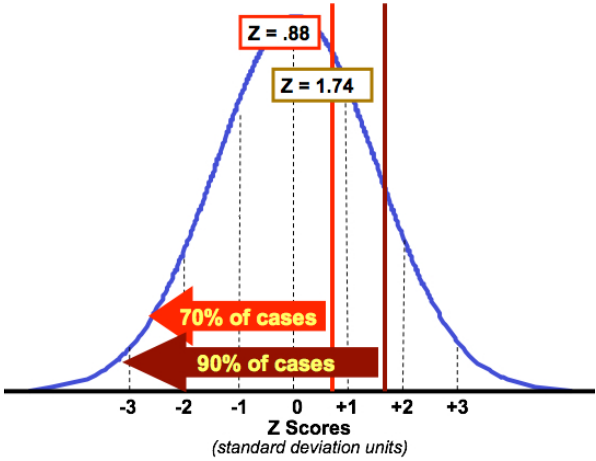
Z is a basic Critical ratio

$$\frac{\text{Distance of the score from the mean} \rightarrow}{\text{Standard Deviation} \rightarrow} = \frac{\text{Strength of the experimental result}}{\text{Error variance or “noise” in the data}}$$

- ✓ In our example the two samples had equally strong scores (X - M).
- ✓ …but differed in the amount of variance in the distribution of scores
- ✓ Weighting the effect – X - M – in each sample by it’s variance [S] yielded different Z scores: .88 v. 1.74.
- ✓ This led us to different judgments of how likely each result would be to have occurred by chance.







Applying the critical ratio to an experiment

$$\text{Critical Ratio} = \frac{\text{Treatment Difference}}{\text{Random Variance (Chance)}}$$

- **In an experiment the Treatment Difference is variance between the experimental and control groups.**
- **Random variance or chance differences among participants within each group.**
- **We evaluate that result by comparing it to a distribution of possible effects.**
- **We *estimate* the distribution of possible effects based on the degrees of freedom ("df").**

We will get to these last 2 points in the next modules.

Examples of Critical Ratios

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

$$\text{Z score} = \frac{\text{Individual Score} - M \text{ for Group}}{\text{Standard Deviation (S) for group}} = \frac{x - M}{s}$$

$$\text{t-test} = \frac{\text{Difference between group } Ms}{\text{Standard Error of the Mean}} = \frac{M_{group1} - M_{group2}}{\sqrt{\frac{Variance_{grp1}}{n_{grp1}} + \frac{Variance_{grp2}}{n_{grp2}}}}$$

$$\text{F ratio} = \frac{\text{Between group differences (differences among} \geq \text{3 group } Ms)}{\text{Within Group differences (random variance among participants within groups)}}$$

$$r \text{ (correlation)} = \frac{\text{Association between variables (joint Z scores) summed across participants } \sum (Z_{variable1} \times Z_{variable2})}{\text{Random variance between participants within variables}}$$

Where would z or t have to fall for you to consider your results "statistically significant"? (Choose a color).

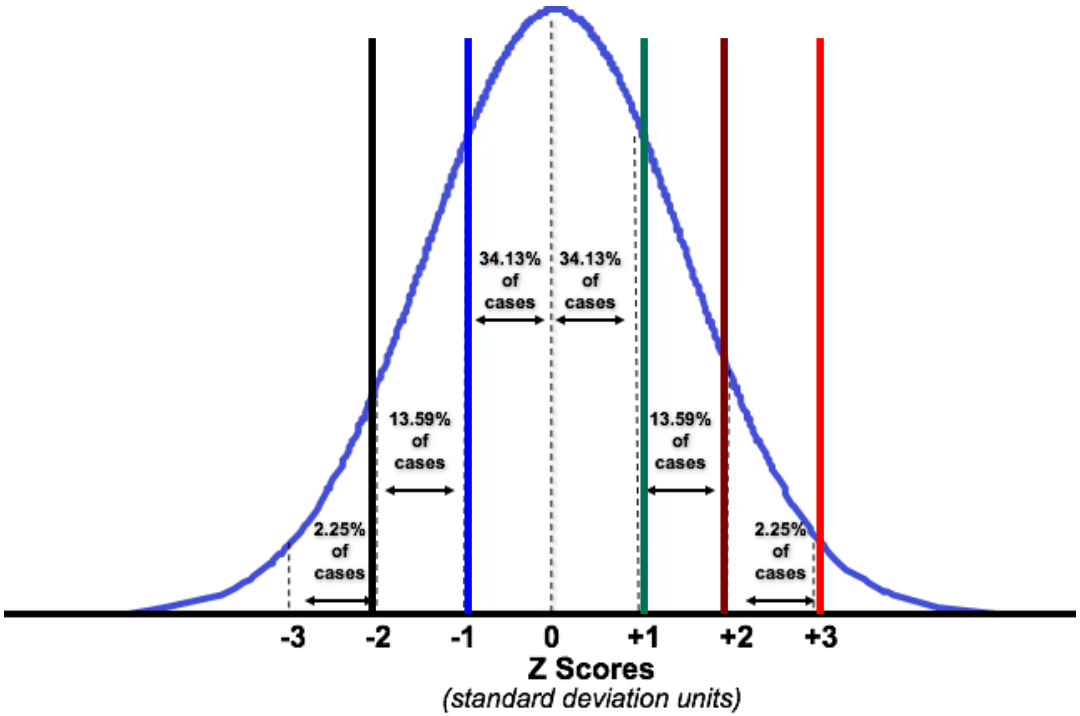
A.

B.

C.

D.

F.



Where would z or t have to fall for you to consider your results "statistically significant"? (Choose a color).

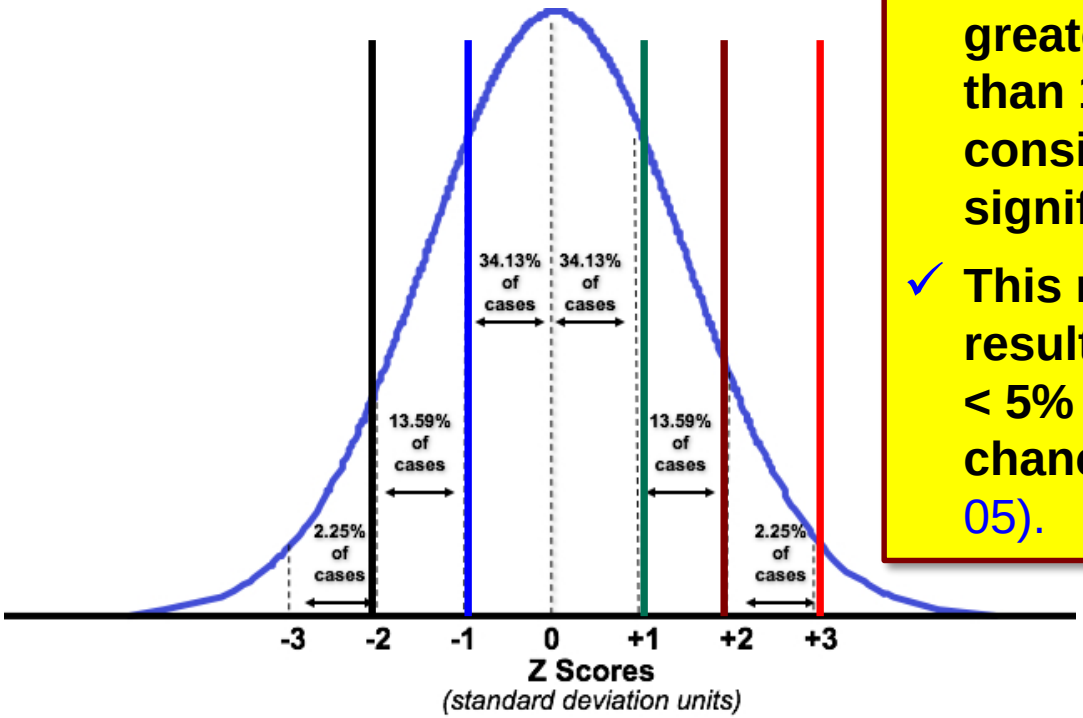
A.

B.

C.

D.

F.



- ✓ **Both of these are correct.**
- ✓ **A Z or t score greater than or less than 1.98 is consided it significant.**
- ✓ **This means that the result would occur < 5% of the time by chance alone** (p < 05).

Where would z or t have to fall for you to consider your results "statistically significant"? (Choose a color).

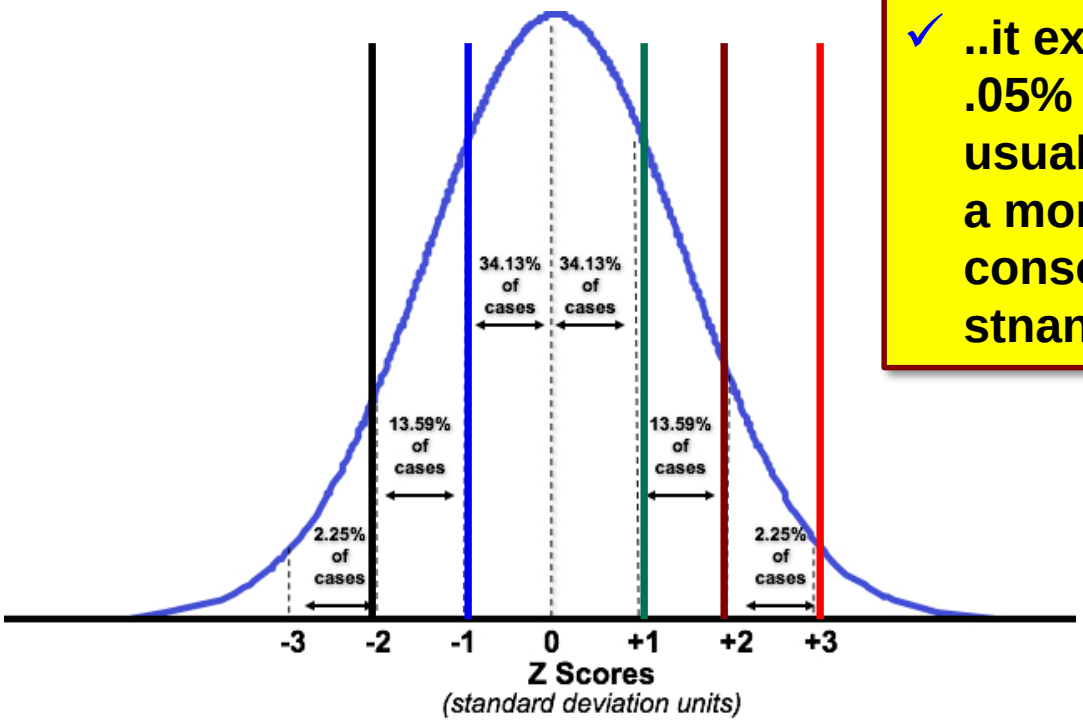
A.

B.

C.

D.

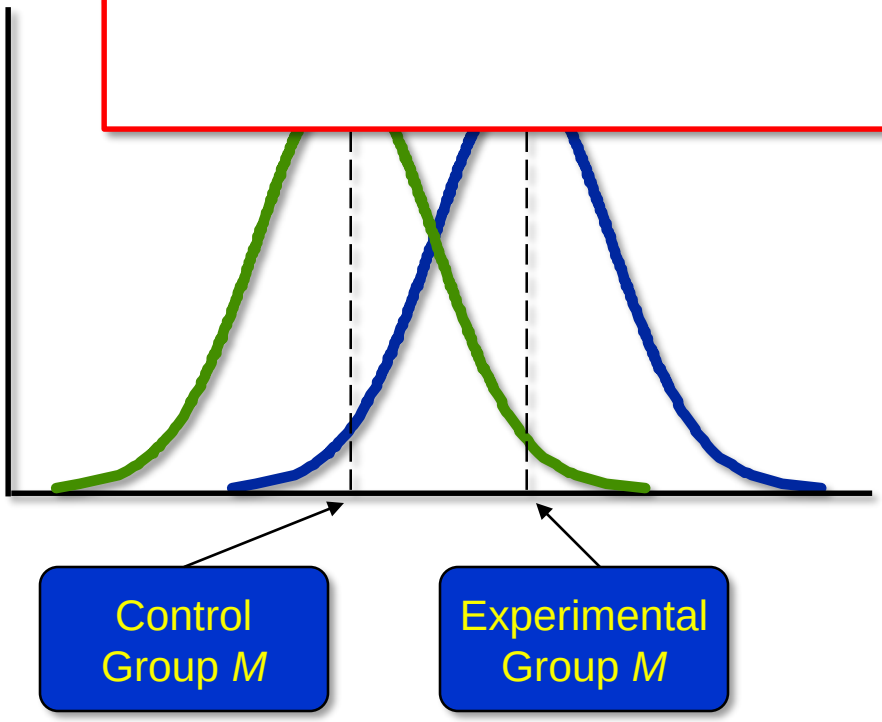
F.



- ✓ **This value would also be statistically significant..**
- ✓ **..it exceeds the .05% value we usually use, so it is a more conservative stnandard.**

t-test: are the *Ms* of two groups *statistically*

Let's apply the critical ratio to an experiment.

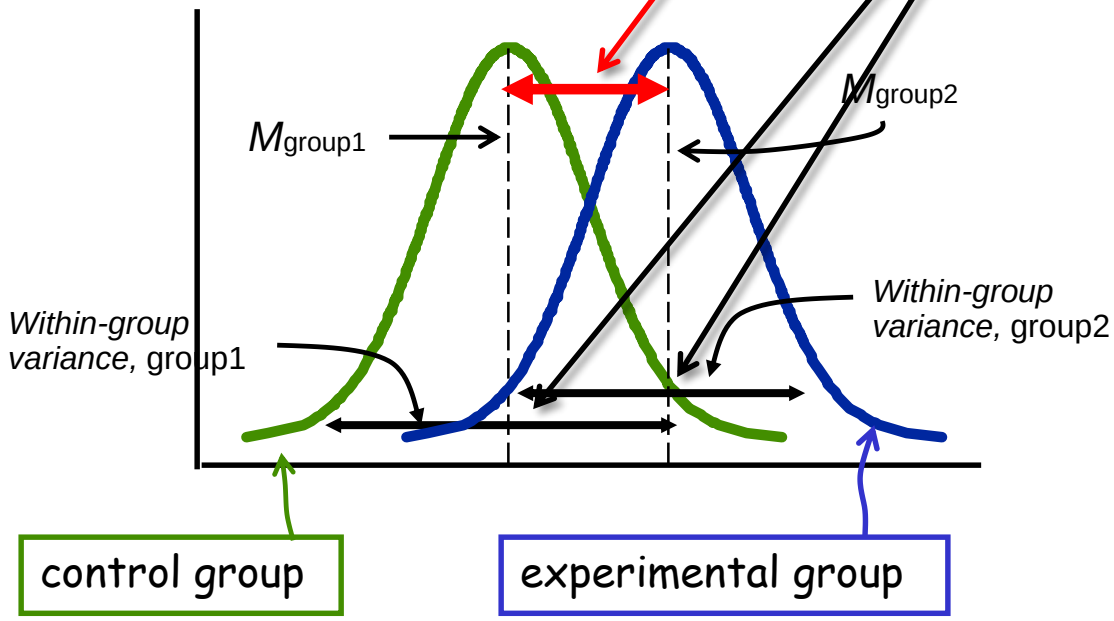


- In **any** experiment the *M*s will differ at least a
- we
- **ly** due to the independent variable.
- Statistically: is the difference between *M*s more than we would expect by **chance alone?**

The critical ratio applied to the t test.

Critical Ratio = D … ps

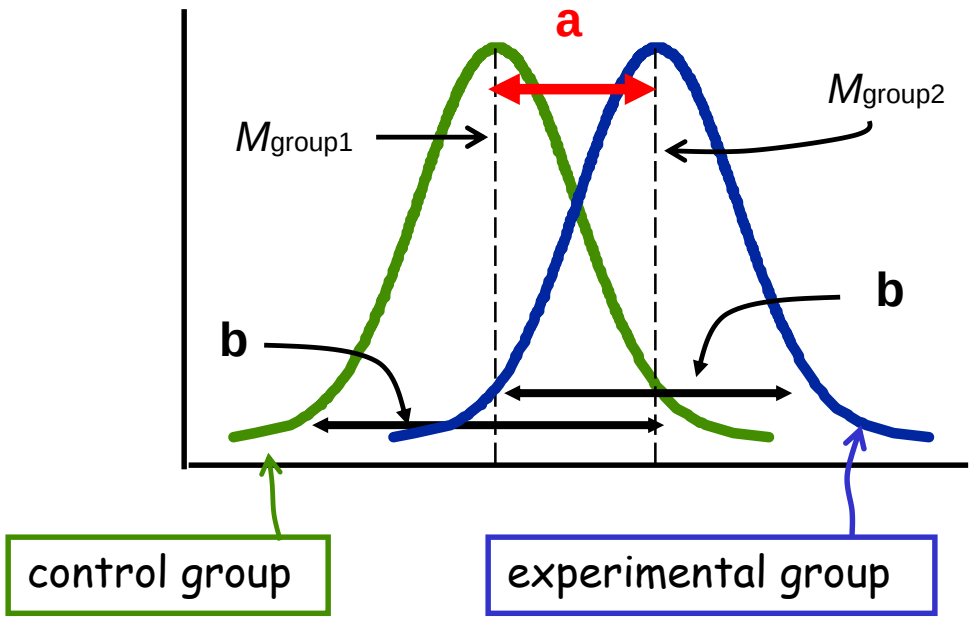
$$\frac{\text{The experimental effect}}{\text{Error variance}}$$



Variance **between groups**

What we would expect by **chance** given the **variance within groups.**

M differences and the Critical Ratio.

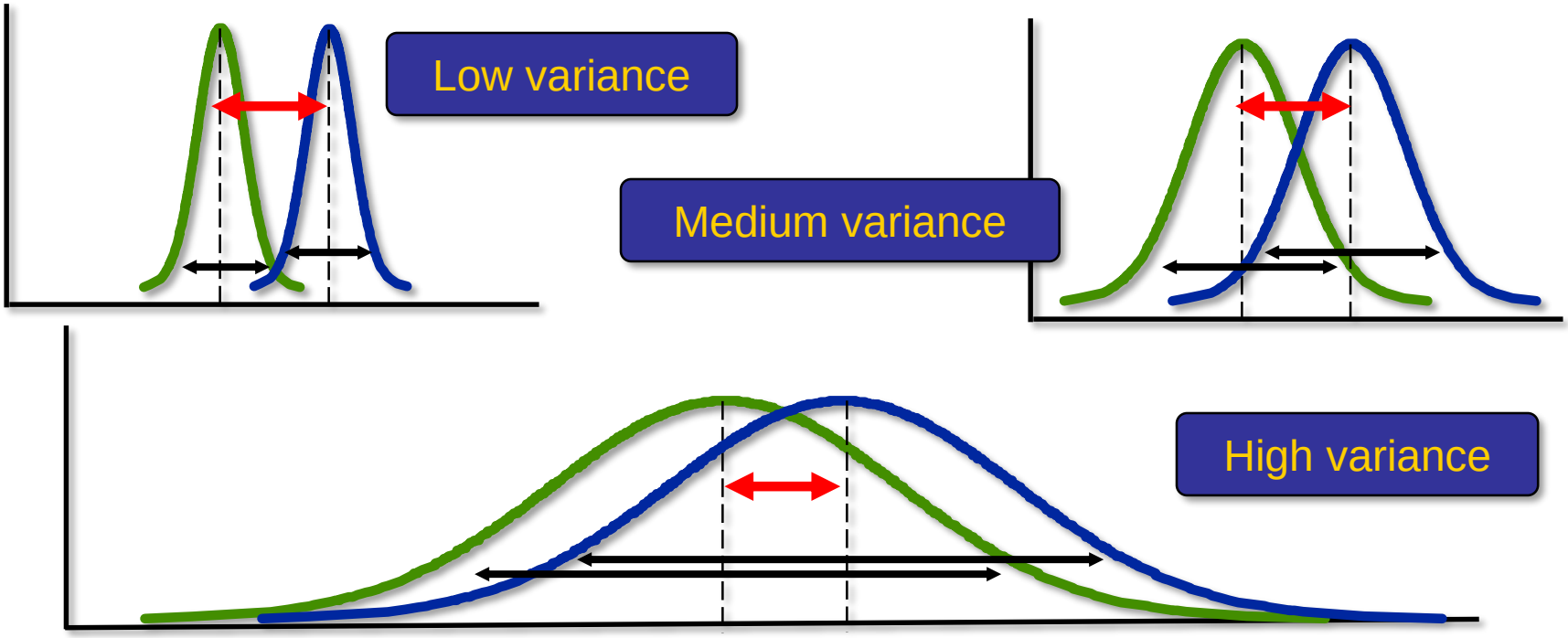


The Critical Ratio in action

All three graphs have = **difference between groups**.

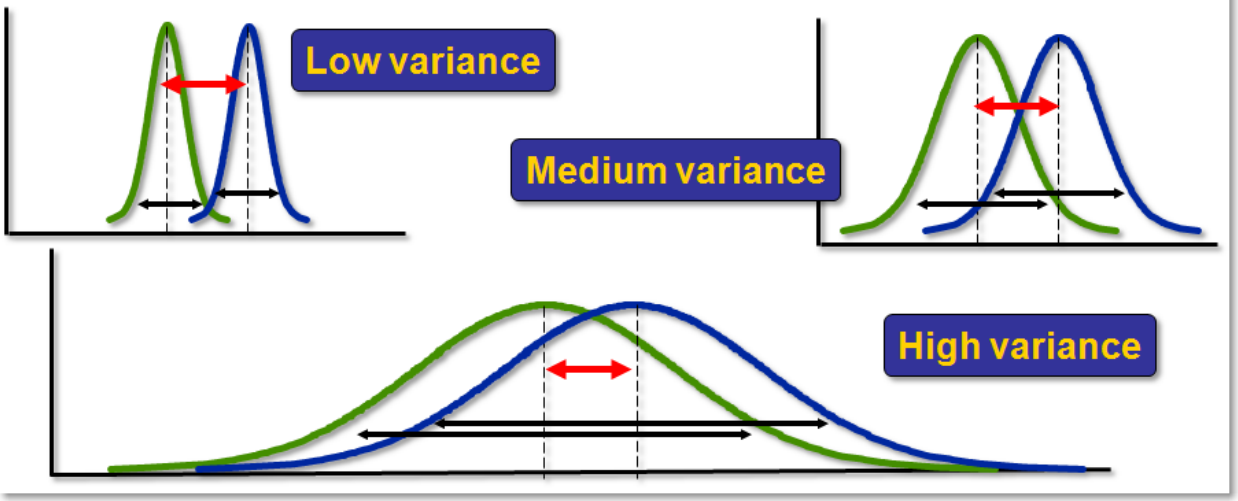
They differ in **variance within groups**.

The critical ratio helps us determine which one(s) represent a **statistically significant** difference.



Clickers!

- Which would the **critical ratio** suggest is a statistically significant difference?



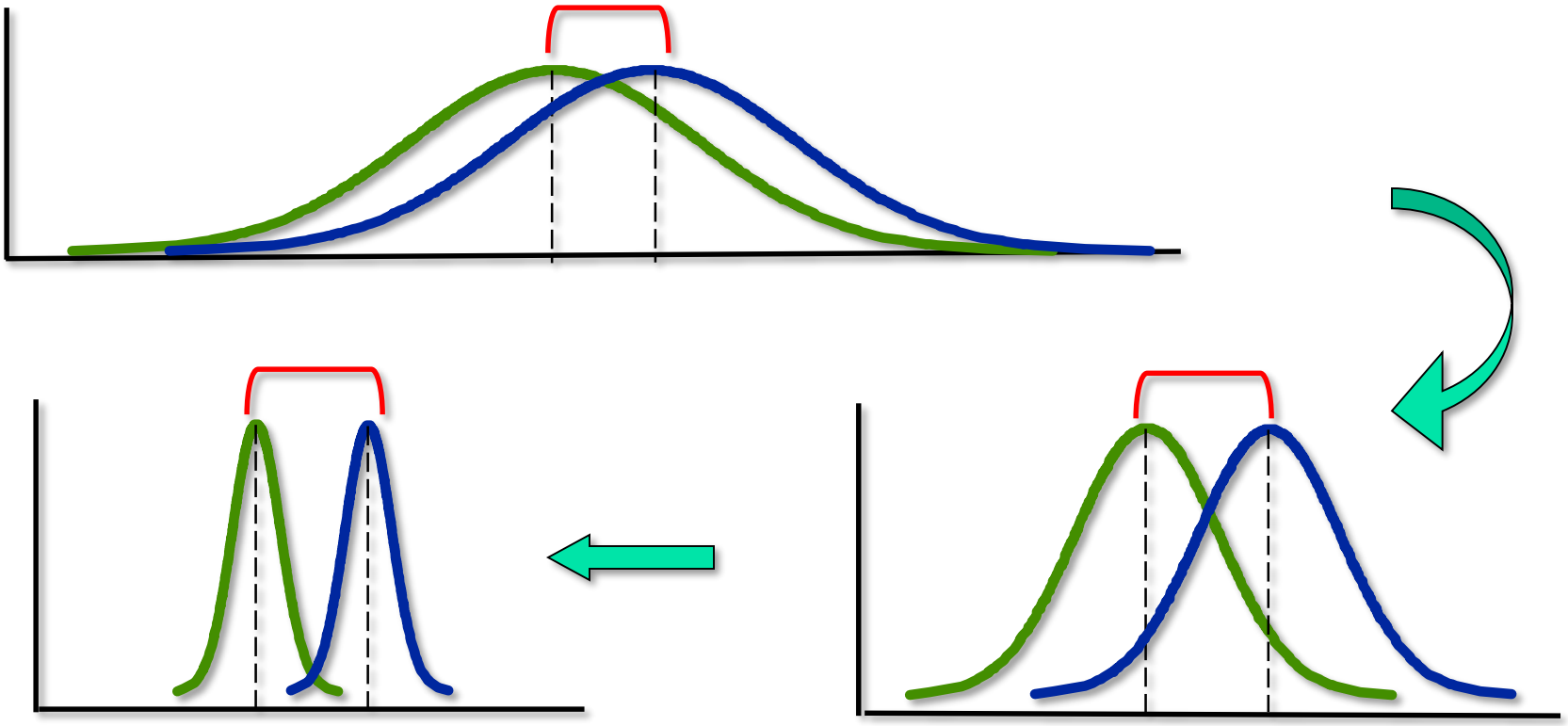
A = All of them
B = Low variance only
C = Medium variance
D = High variance
E = None of them

Critical ratio and variances, 1

Critical ratio:

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

Gets larger as the variance(s) decreases, given the same M difference…..

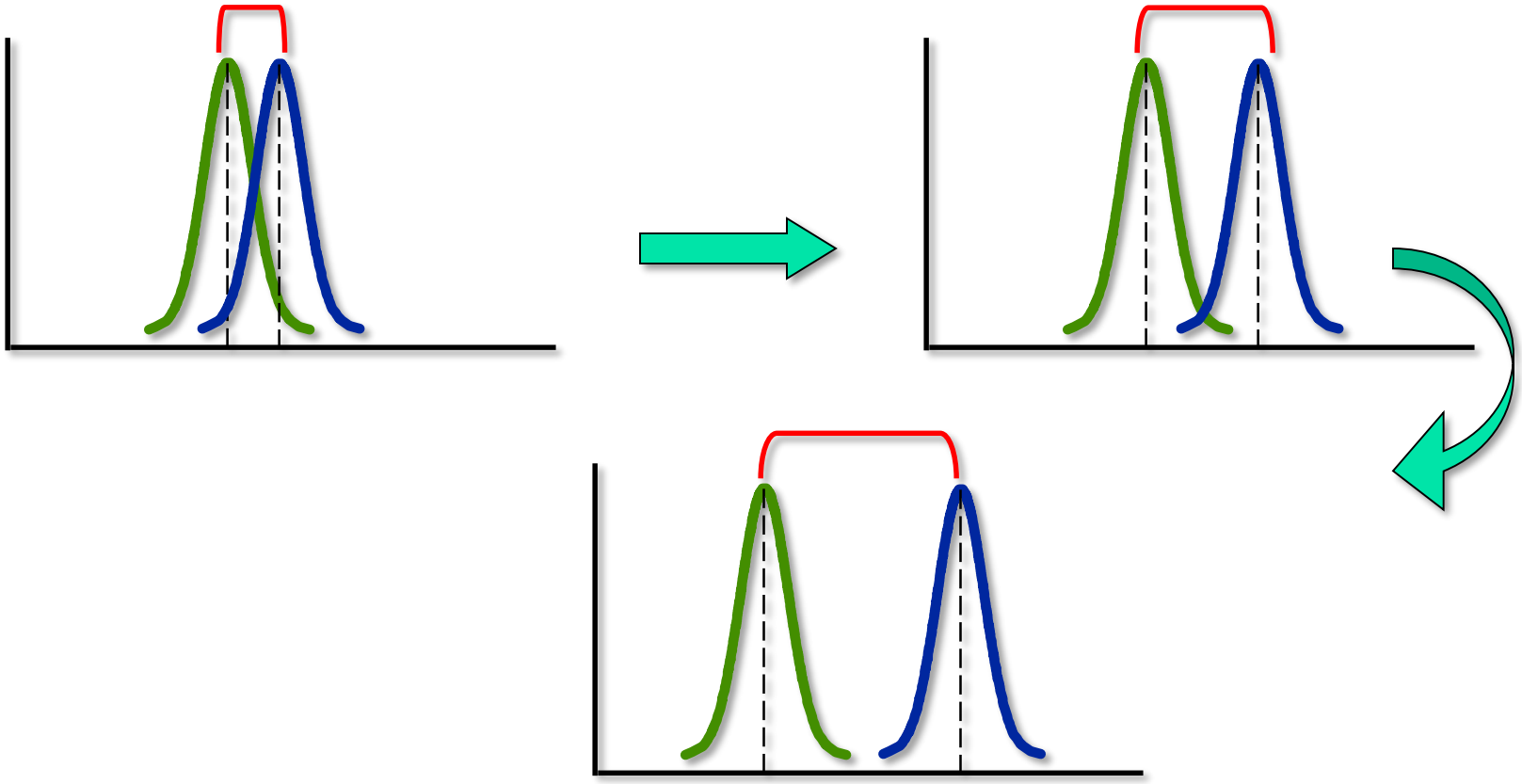


Critical ratio and variances, 2

Critical ratio:

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

…also gets larger as the M difference increases, even with same variance(s)

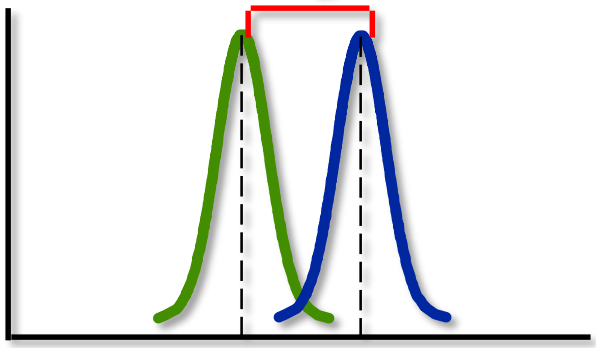


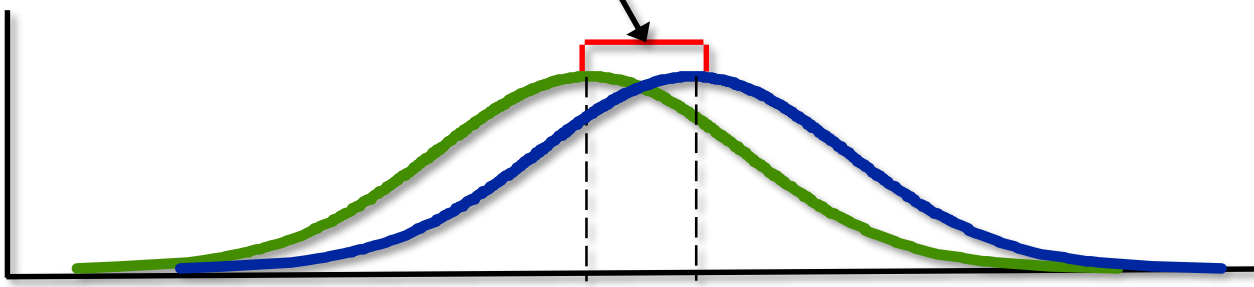
What Do We Estimate; experimental effect

$$\frac{\text{Experimental Effect}}{\text{Error variance}}$$

Difference between group Ms

M difference (between control & experimental groups) is the same in both data sets



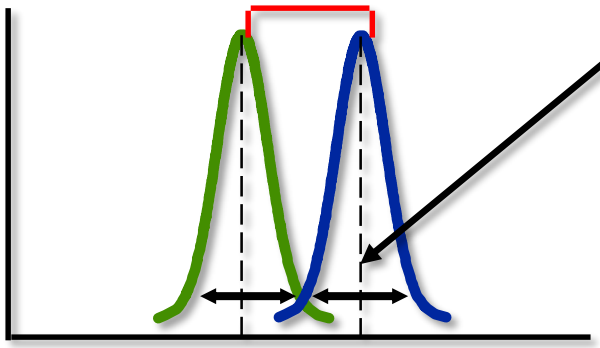


What Do We Estimate: error term

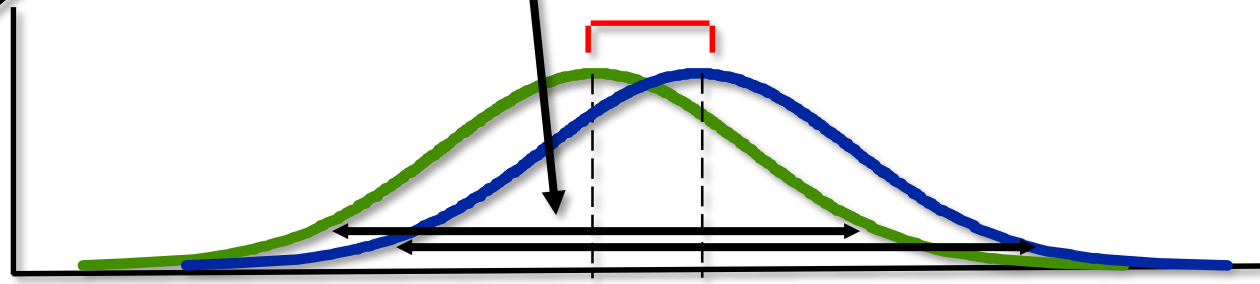
$$\frac{\text{Experimental Effect}}{\text{Error variance}}$$

Variability within groups

Variances differ a lot in the two examples



Low variability

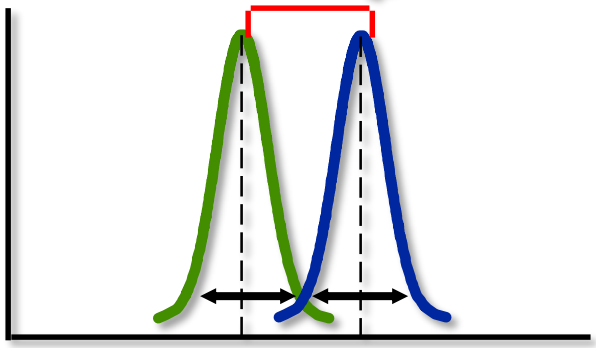


High variability

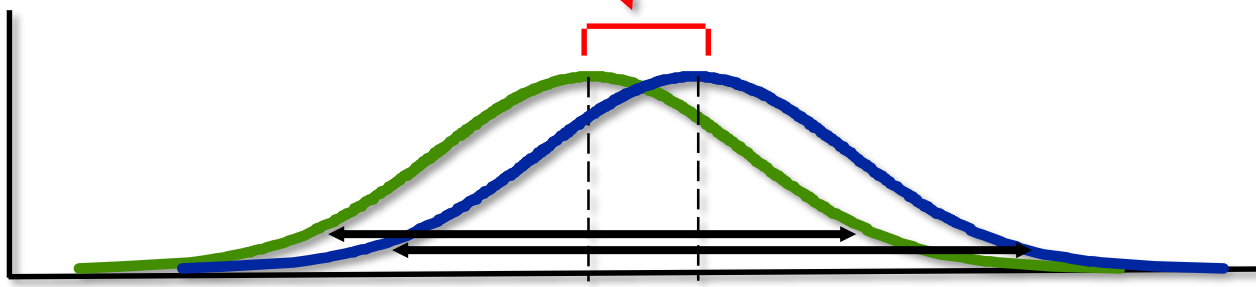
Assigning numbers to the critical ratio: numerator

$$\frac{\text{Experimental Effect}}{\text{Error variance}} = \frac{\text{Difference between group Ms}}{\text{Variability within groups}}$$

$$= \frac{(M_{group1} - M_{group2}) - 0}{}$$



Low variability

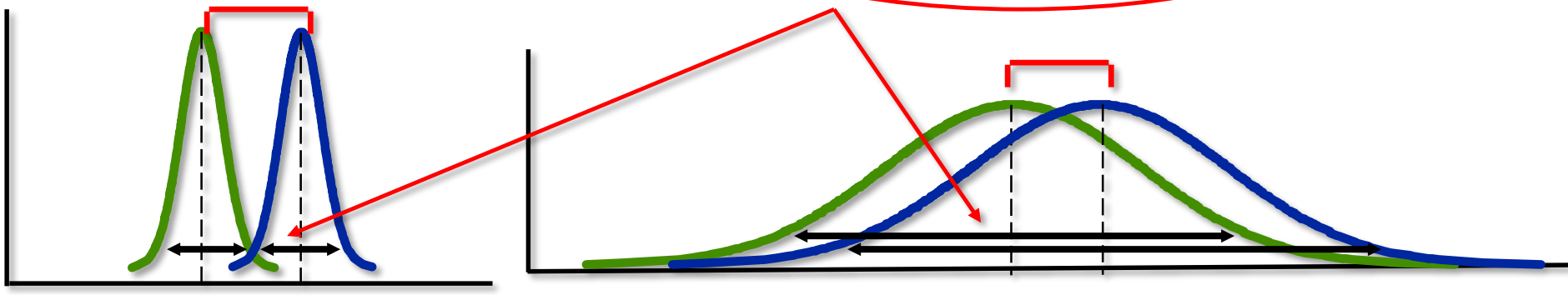


High variability

Assigning numbers to the critical ratio: denominator

$$\frac{\text{Experimental Effect}}{\text{Error variance}} = \frac{\text{Difference between Ms}}{\text{Variability within groups}}$$

$$= \frac{M_{group1} - M_{group2}}{\text{Standard error:}\ \sqrt{\frac{\text{Variance}_{grp1}}{n_{grp1}} + \frac{\text{Variance}_{grp2}}{n_{grp2}}}}$$



Low variability

High variability

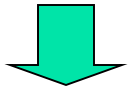
Crit

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

- Experimental effect “adjusted” by the variance.
- Yields a score: Z, t, *r,* etc.
  - Positive: grp1 > grp2
  - …or Negative: grp1 < grp2.
- Any critical ratio [CR] is likely to differ from 0 by chance alone.
  - Even in “junk” data two groups may differ.
  - Cannot simply test whether Z or t is “absolutely” different than 0.
- We evaluate whether the CR is greater than what we expect *by chance alone*.

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

- **A large CR is likely <u>not</u> due only chance – it probably reflects a "real" experimental effect.**
  - The difference *between* groups is very large relative to the error (*within*-group) variance
- **A very small CR is almost certainly just error.**
  - Any difference between groups is not distinguishable from error or simple chance: group differences may not be due to the experimental condition (Independent Variable).
- **A mid-size CR? How large must it be to assume it did not occur just by chance?**



We answer this by comparing it to a (hypothetical) distribution of *possible* CRs.

Distributions of Critical Ratios

➔ Imagine you perform the same experiment 100 times.

- ✓ You randomly select a group of people
- ✓ You randomly assign ½ to the experimental group, ½ to control group
- ✓ You run the experiment, get the data, and analyze it using the critical ratio:

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

$$= \frac{M_{group1} - M_{group2}}{\sqrt{\frac{Variance_{grp1}}{n_{grp1}} + \frac{Variance_{grp2}}{n_{grp2}}}} = t$$

➔ Imagine you perform the same experiment 100 times.

- ✓ Then … You do the same experiment again, with another random sample of people…
- ✓ And get a critical ratio (t score) for those results…

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

$$= \frac{M_{group1} - M_{group2}}{\sqrt{\frac{Variance_{grp1}}{n_{grp1}} + \frac{Variance_{grp2}}{n_{grp2}}}} = t$$

Distributions of Critical Ratios

➔ Imagine you perform the same experiment 100 times.

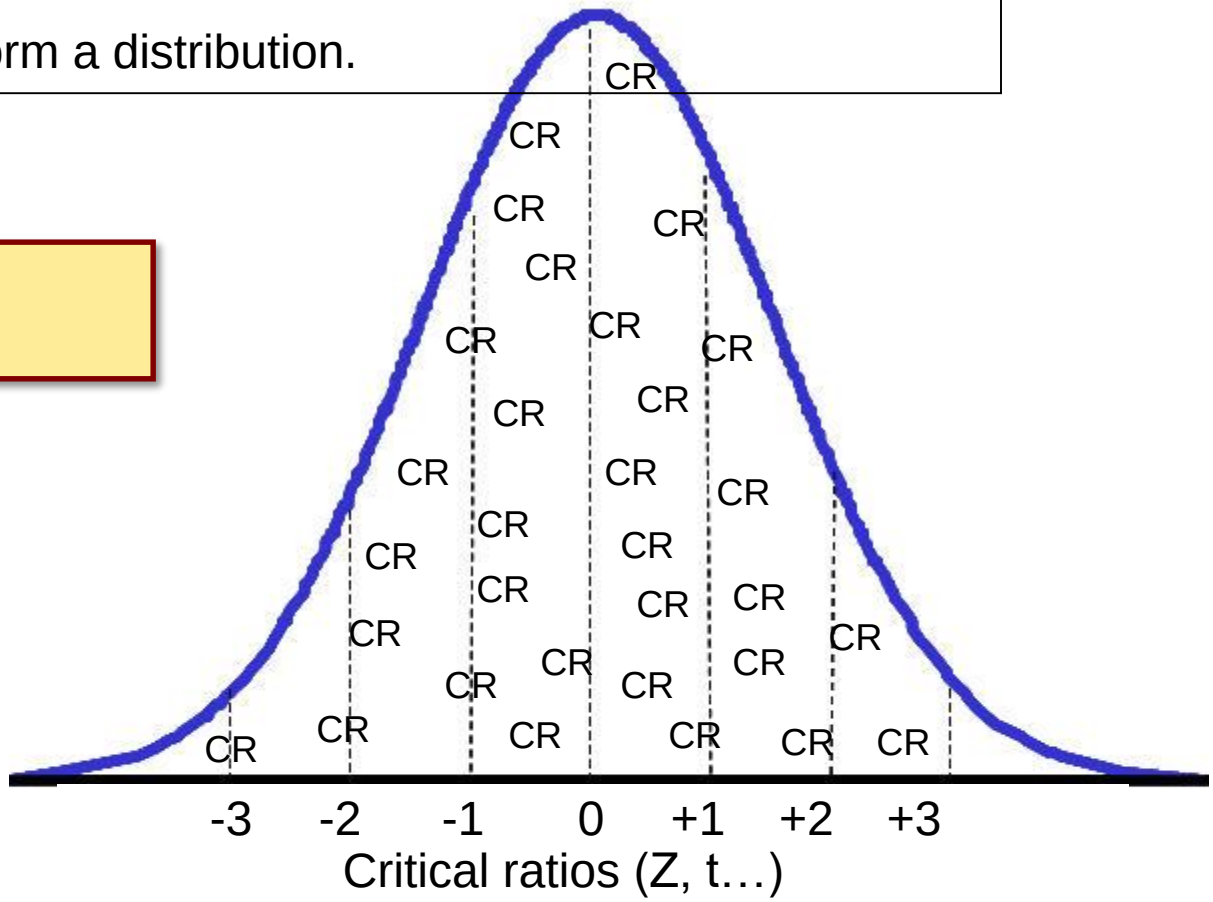
- ✓ And you get yet another sample…
- ✓ And get a critical ratio (t score) for those results…

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

$$= \frac{M_{group1} - M_{group2}}{\sqrt{\frac{Variance_{grp1}}{n_{grp1}} + \frac{Variance_{grp2}}{n_{grp2}}}} = t$$

- ➔ Each time you (hypothetically) run the experiment you generate a critical ratio (CR).
  - ✓ For a simple 2-group experiment the CR is a t ratio
  - ✓ It could just as easily be a Z score, an *F* ratio, an *r*…
- ➔ These **Critical Ratios** form a distribution.

This is called a **Sampling Distribution.**



- ➔ Imagine you perform the same experiment 100 times.
- ➔ Each experiment generates a critical ratio [Z score, t ratio…]
- ➔ These Critical Ratios form a distribution

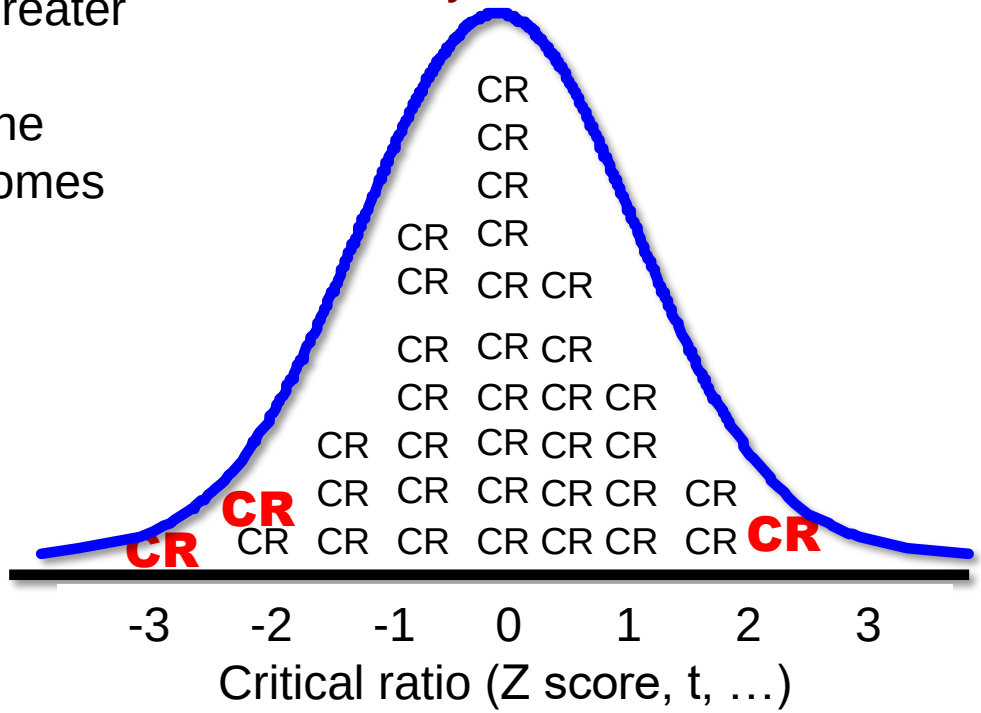
This is called a **Sampling Distribution.**

Most Critical Ratios will cluster around '0'

- ✓ M = 0
- ✓ Progressively fewer are greater or less than 0.
- ✓ With more observations the sampling distribution becomes "normal"

Null hypothesis; there is no real effect, so any CR above or below 0 is by chance alone.

More extreme scores are unlikely to occur by chance alone.


CR
CR
CR
CR CR
CR CR CR
CR CR CR
CR CR CR CR
CR CR CR CR CR
CR CR CR CR CR CR
CR CR CR CR CR CR CR
CR CR CR CR CR CR CR CR
-3 -2 -1 0 1 2 3
Critical ratio (Z score, t, …)

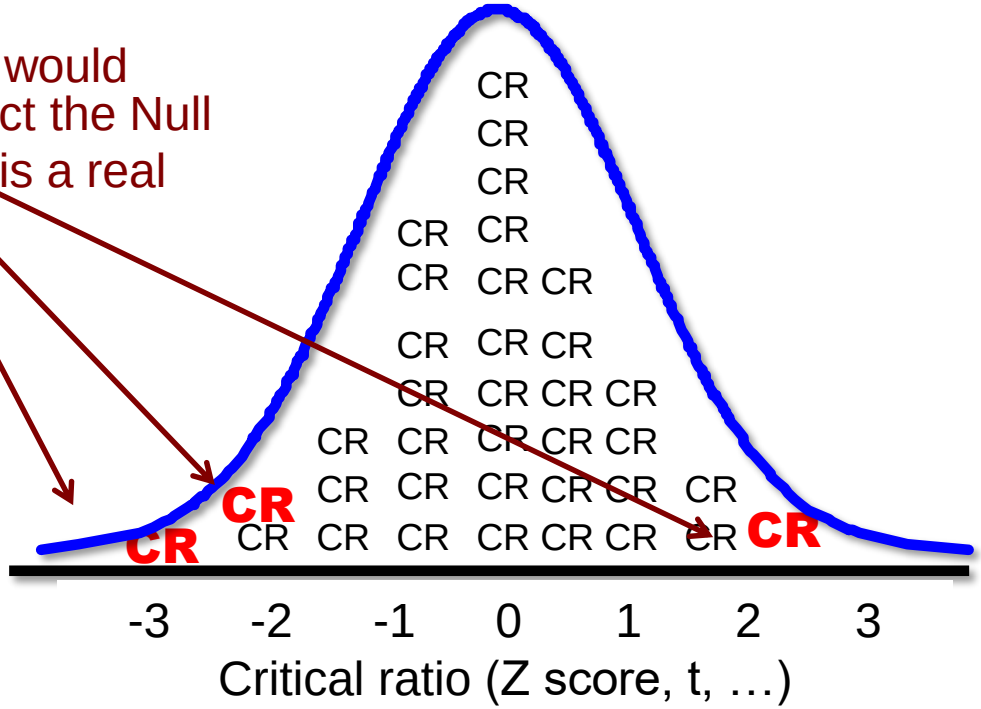
- Imagine you perform the same experiment 100 times.
- Each experiment generates a critical ratio [Z score, t ratio…]
- These **Critical Ratios** form a distribution

This is called a **Sampling Distribution.**

Most Critical Ratios will cluster around '0'
(M = 0)

If a critical ratio is **larger** than we would expect by chance alone, we Reject the Null hypothesis and accept that there is a real effect.

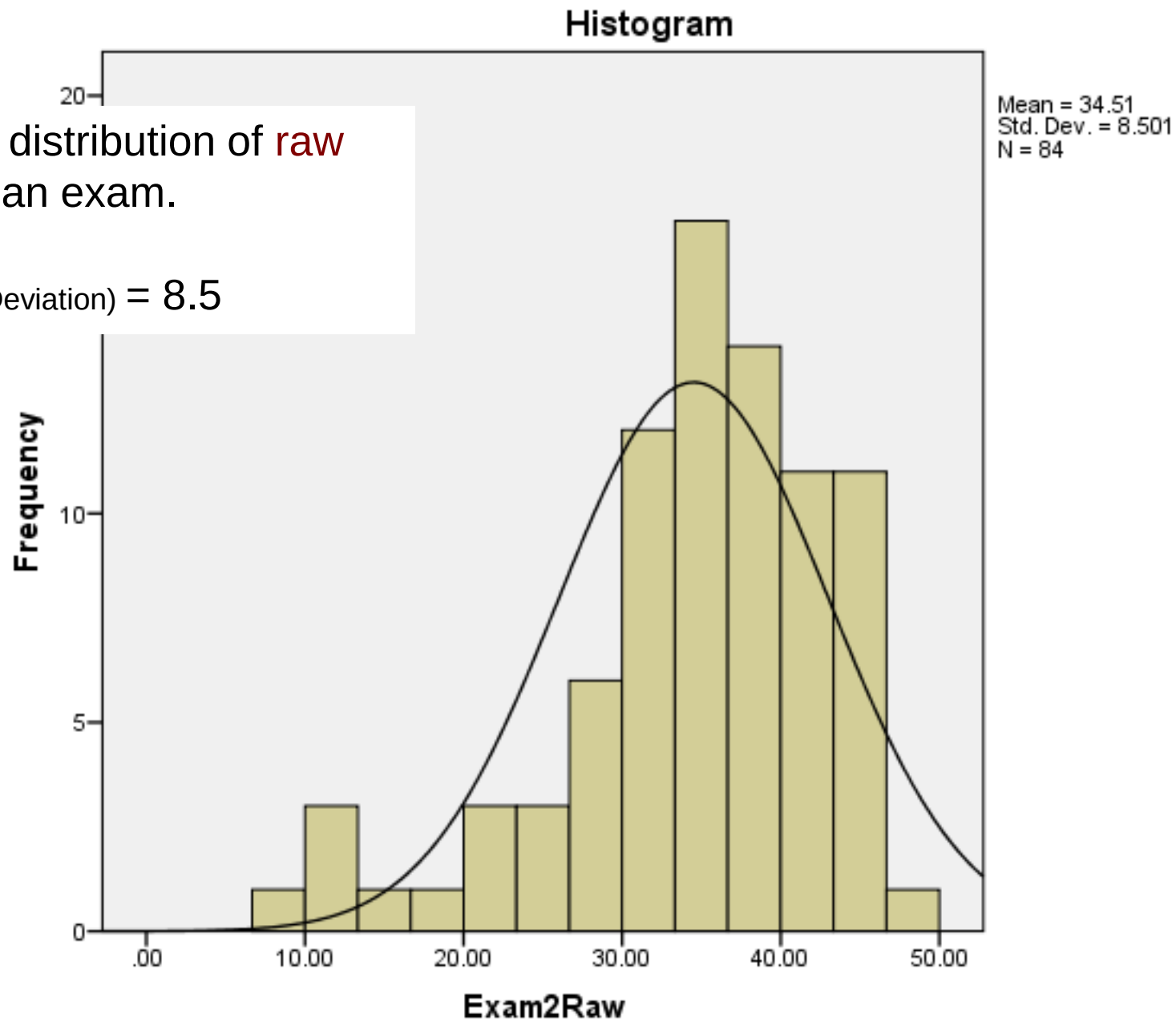
More extreme scores are unlikely to occur by chance alone.



This is the distribution of raw scores for an exam.

M =34.5

S (Standard Deviation) = 8.5



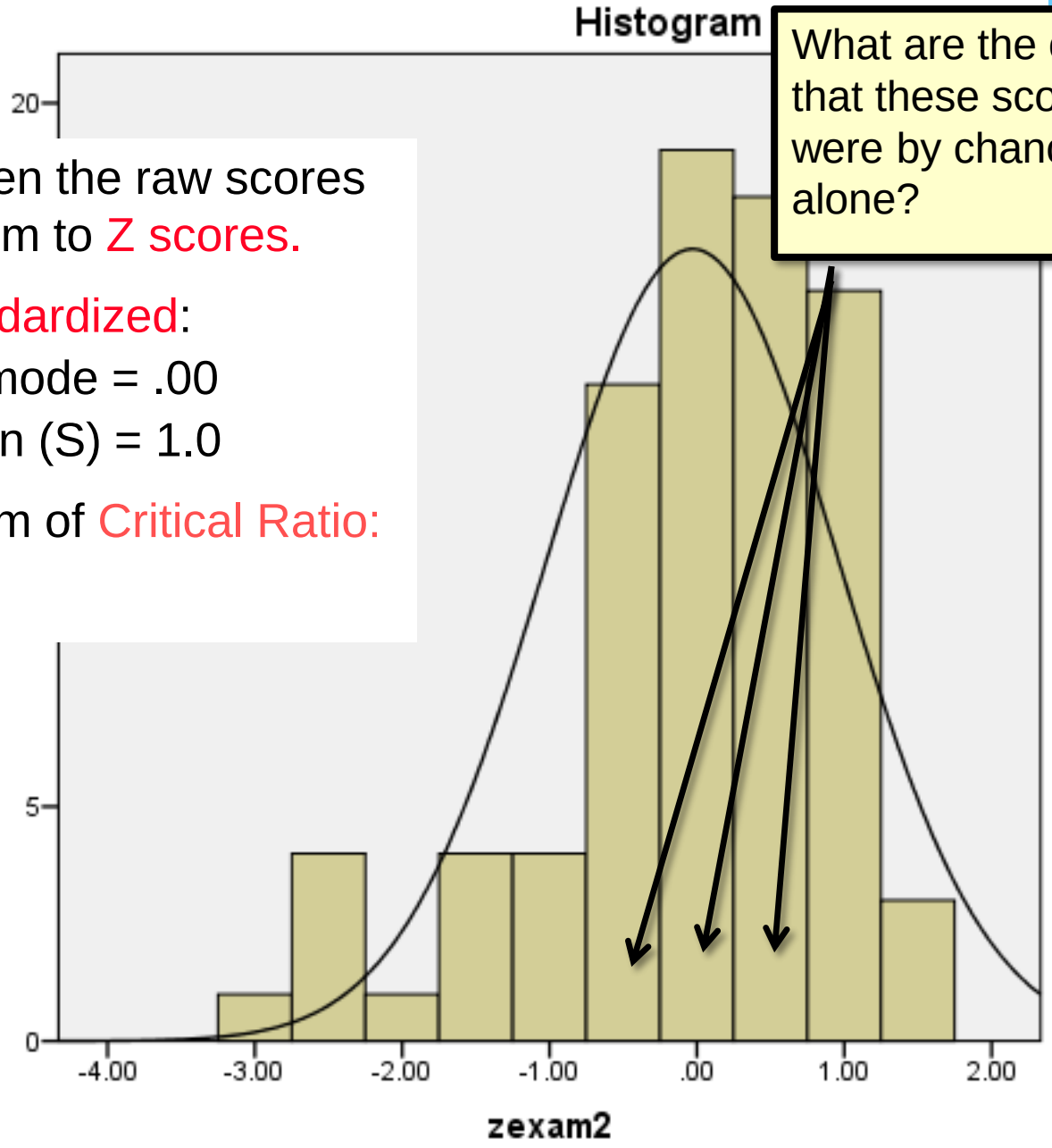
Here we have taken the raw scores and converted them to Z scores.

Z scores are Standardized:

Mean, median & mode = .00

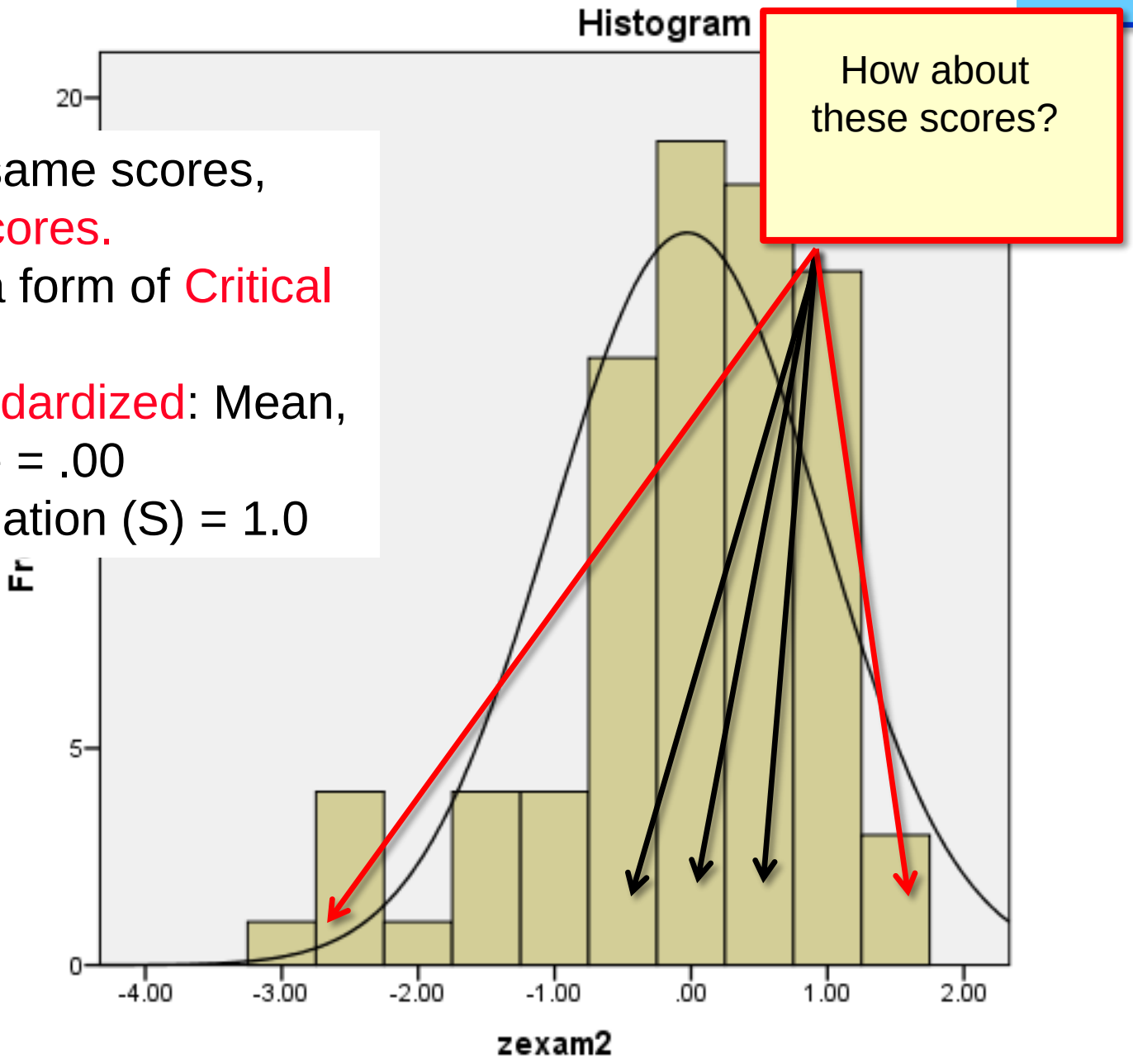
Standard Deviation (S) = 1.0

Z scores are a form of Critical Ratio:



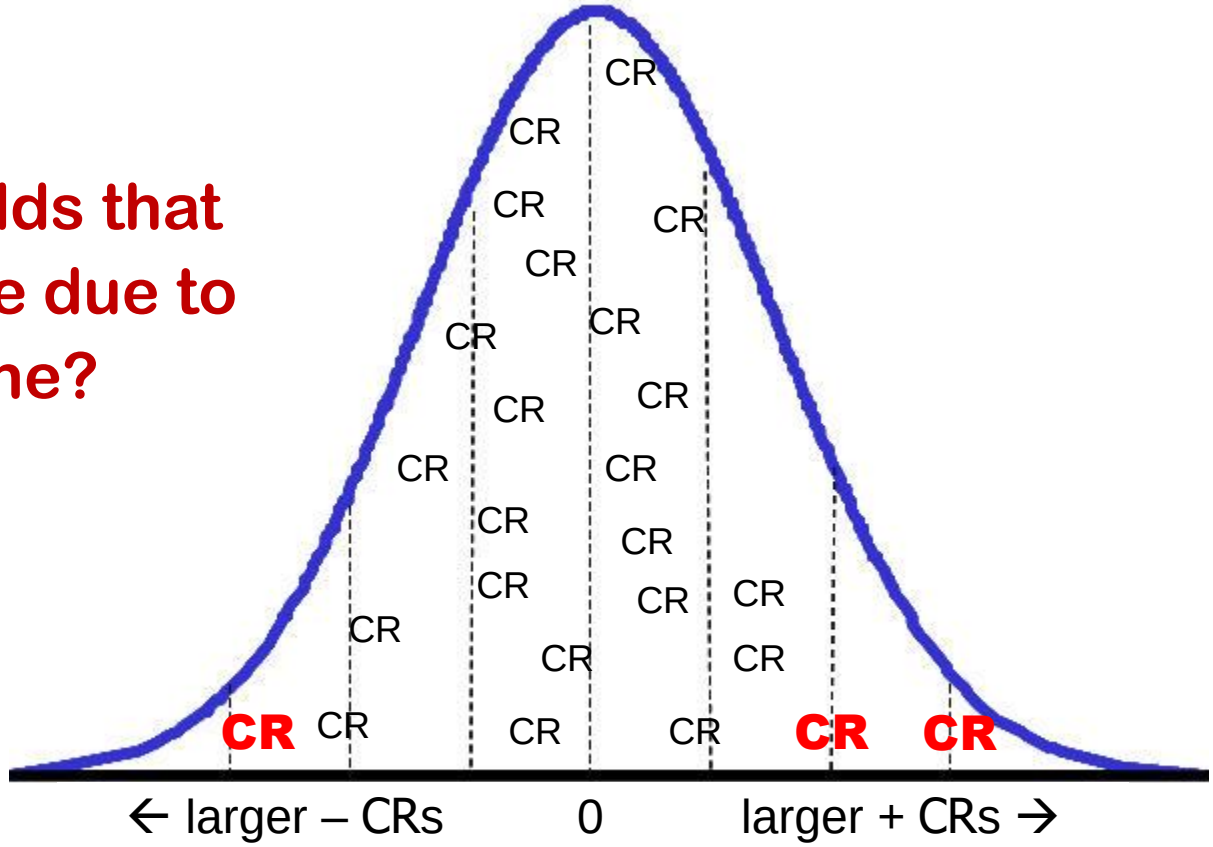
What are the odds that these scores were by chance alone?

Here are the same scores, shown as Z scores.
Z scores are a form of Critical Ratio
They are Standardized: Mean, median, mode = .00
Standard Deviation (S) = 1.0



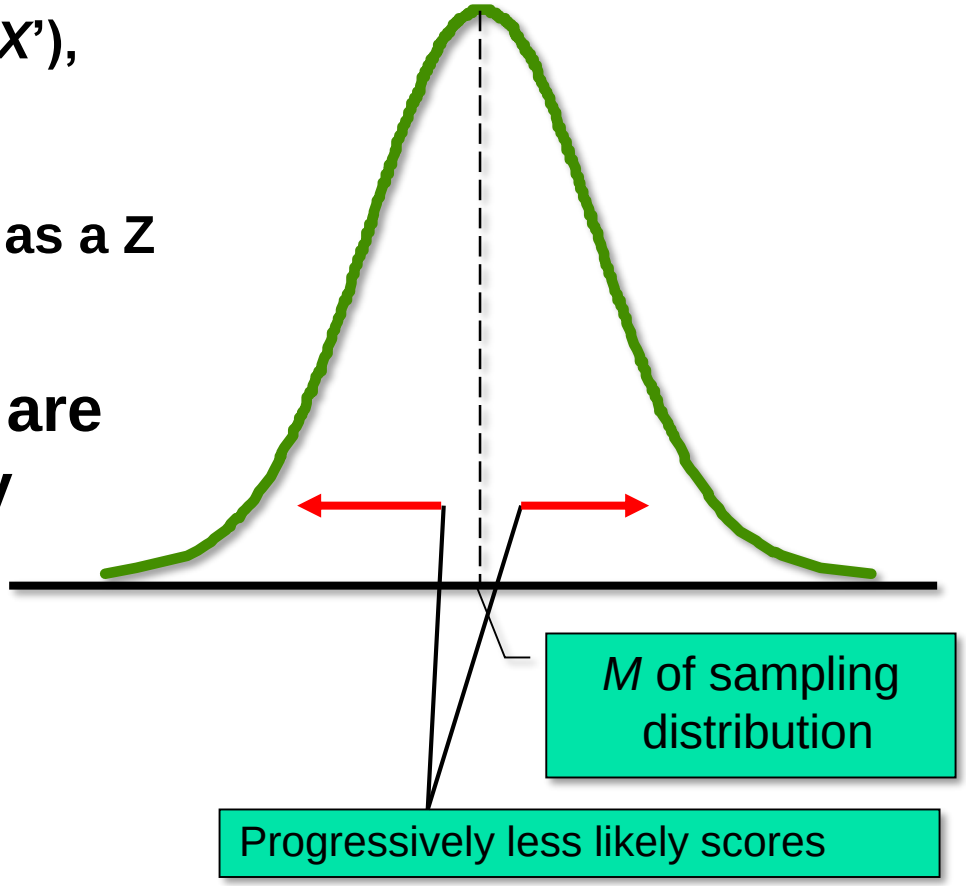
After we conduct our experiment and get a result (a critical ratio or t score) our question is…

What are the odds that these results are due to chance alone?


CR
CR
CR

We infer *statistical significance* by locating a score along the normal distribution.

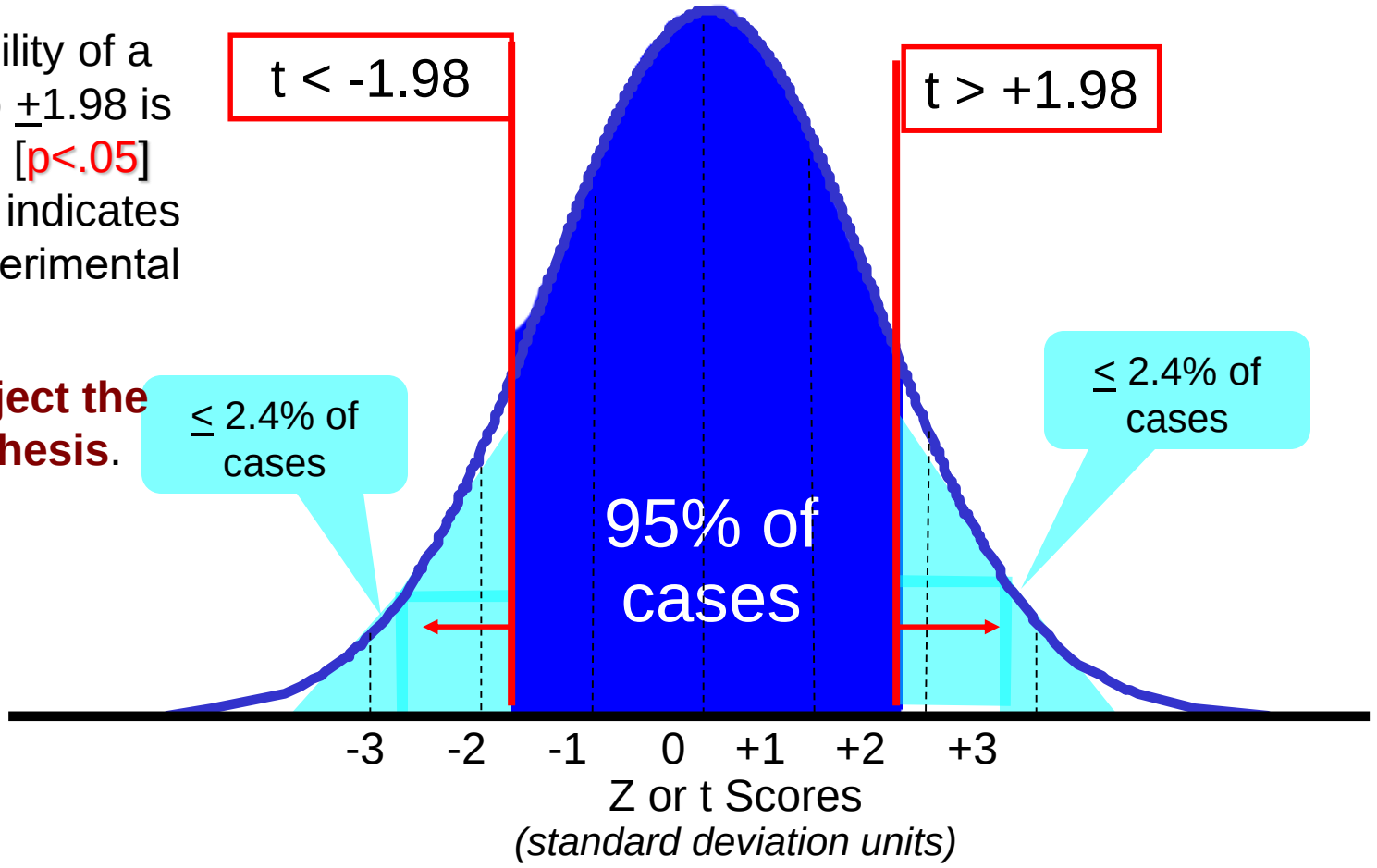
- **A *score* can be:**
  - **An individual score (‘*X*’),**
  - **A group *M*,**
  - **A Critical Ratio such as a Z or t score.**
- **More extreme scores are less likely to occur by chance alone.**



- A Z or t score that exceeds ± 1.98 would occur *by chance alone* less than 5% of the time.

The probability of a critical ratio ±1.98 is low enough [p<.05] that it likely indicates a "real" experimental effect.

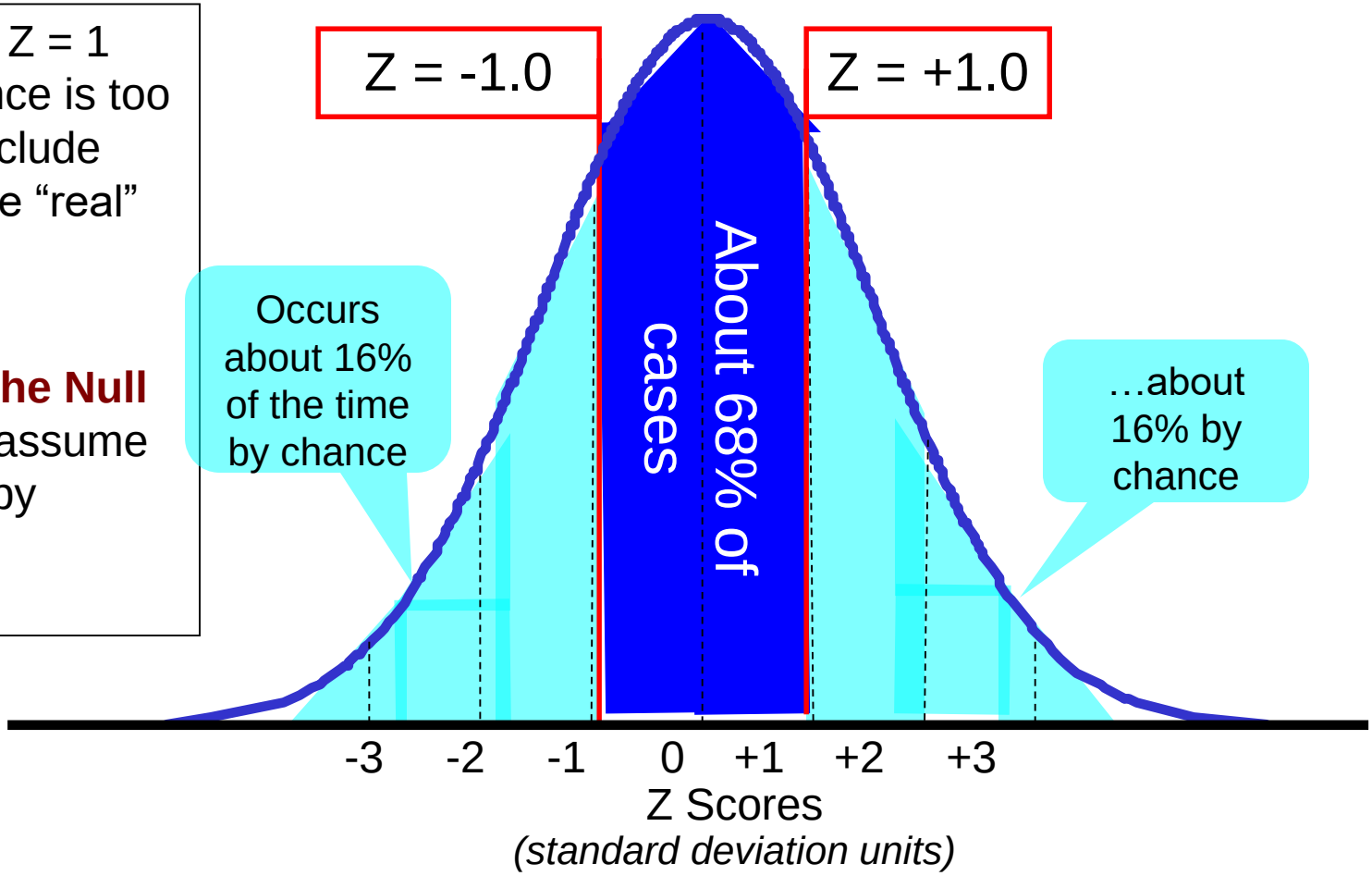
We then **reject the Null Hypothesis**.



- A Z or t score that exceeds ± 1.0 would occur *by chance alone* about 32% of the time.

The probability of Z = 1 occurring by chance is too high for us to conclude that the results are "real" (i.e., "statistically significant").

We then **accept the Null Hypothesis** and assume that any effect is by chance alone.



Summary

✓ **Statistical decisions follow the critical ratio:**

$$\text{Critical ratio} = \frac{\text{Experimental effect}}{\text{Error variance}}$$

✓ **Z is the prototype critical ratio:**

$$Z = \frac{\text{Distance of the score } (X) \text{ from the mean } (M)}{\text{Variance among all the scores in the sample [standard deviation } (S)]} = \frac{X - M}{S}$$

✓ **t is also a basic critical ratio** used for comparing groups:

$$t = \frac{\text{Difference between group Means}}{\text{Variance within the two groups [standard error of the M } (SE)]} = \frac{M_1 - M_2}{\sqrt{\frac{\text{Variance}_{grp1}}{n_{grp1}} + \frac{\text{Variance}_{grp2}}{n_{grp2}}}}$$

The critical ratio

The next modules will:

- **discuss the logic of scientific** (statistical) **reasoning,**
- **show you how to derive a t value.**